February '19
III Edition

GEN(E)IUS

The Student-Run Bioinformatics Magazine

Are Dreams Inherited?

Are dreams connected to genetics and inheritance? Find out!

Baseball Statistics meet Computational Biology

An interesting article on how Biology meets baseball.

Computational Biology Group
@SASTRA University
www.sastra.edu/combigs

Contents

Editors: Athirstaa Sai Prashath
Sowmya Shivathmika

Design team: Angelin Jenefer Keerthi Bavadharani

Cover Design: Tamilarasi

GEN(E)IUS

The COMBIGS Magazine III Edition

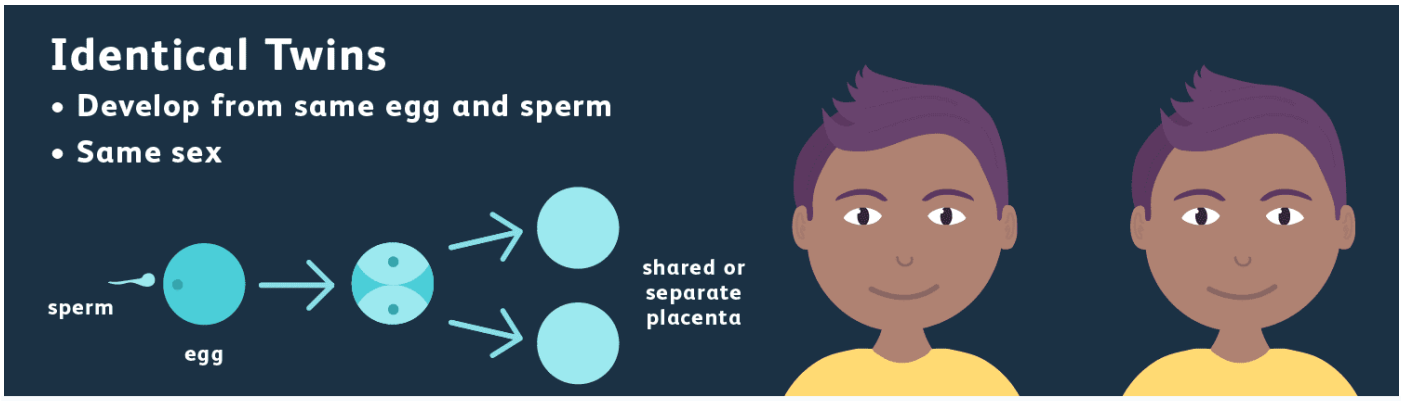
FOREWORD

COMBIGS, the student run committee by the department of Bioinformatics is delighted to launch the third edition of GEN(E)IUS, a student run magazine.

COMBIGS was started in the year 2005, with the main aim of creating awareness and exposing the students to the latest developments in Bioinformatics. This is achieved by organizing symposiums, conferences, publishing a bi-semester magazine, a well-connected alumni network and several such works.

In 2012, COMBIGS organized the first International Conference on Structural and Functional Genomics. In 2016, COMBIGS was happy to release its first edition of the e-magazine at the 2nd International Conference on Structural and Functional Genomics. In 2018, the International Conference on Genome Biology and Health Sciences was conducted. COMBIGS also has a bioinformatics library which contains journals, magazines, reference books and access to several online resources set up in the department.

A subsidiary group of COMBIGS called the BIOINFORUM was created in the year 2018, with special emphasis on educating and creating awareness among the undergraduate students. BIOINFORUM conducts several activities such as weekly meetings, competitions, online coding contests, debates, and interactive sessions with a student guest speaker, all of which have seen great success.



TWINS: THEIR INDIVIDUALITY INSIGHTS!

INTRODUCTION:

Twins and studies associated with them, remain the best tool for the behavioral genetics. They have a special claim upon everyone's attention. But they do have their own perspective personalities. To compute the inheritance of traits with respect to variations, modern twin studies are trying to assess the effect of an individual's shared environment (family) and unique environment (the characteristic events that configure a life) on a trait.

PUBLIC PERCEPTION ON GENE MECHANISMS:

- Generally, identical twins (monozygotic) inherit all their genes from their parents, while fraternal twins (dizygotic), share only the 50% of the genes.
- Traits can be inherited. For each trait governed by dominant genetic mechanisms, a dominant gene inherited from one of the parent trumps a recessive gene inherited from the other. For example, if a person inherits a recessive gene for brown hair from one parent and a dominant gene for black hair from the other, then the dominant black gene is expressed and the offspring's hair colour becomes black.
- Mostly, we are familiar with additive gene mechanism. i.e., we will be having an interesting characteristic. For the above example, we will be having blackish brown hair.
- Twin studies - along with emerging genome and molecular research methods.
- Twin studies only focus on how the trait is inherited. Besides, Molecular genetics attempts to pinpoint the affects the particular gene that references to the contribution to overall pattern of inheritance.

IDEOLOGY:

Identical twins (monozygotic) develop from a single fertilized egg and have the same genome. Any difference between the twins is due to their environment not genome.

Identical twins have their own genetics inherited from their parents, but then they differ in their actions. It mostly depends on the experiences they face and the situations they handle. Although they are raised from the same household and environment, they behave differently, because of their individual experiences that plays a vital role. The neurons from the brain promote these behavioral patterns. Different experiences lead to epigenetic changes and consequently, modifications in the phenotype as well.

CONCLUSION:

Genes find the pattern characteristics and finally the individual micro-environment shape their identity. Distinct environment, parenting and education also reflect in epigenetics that affects the cells reading the genome.

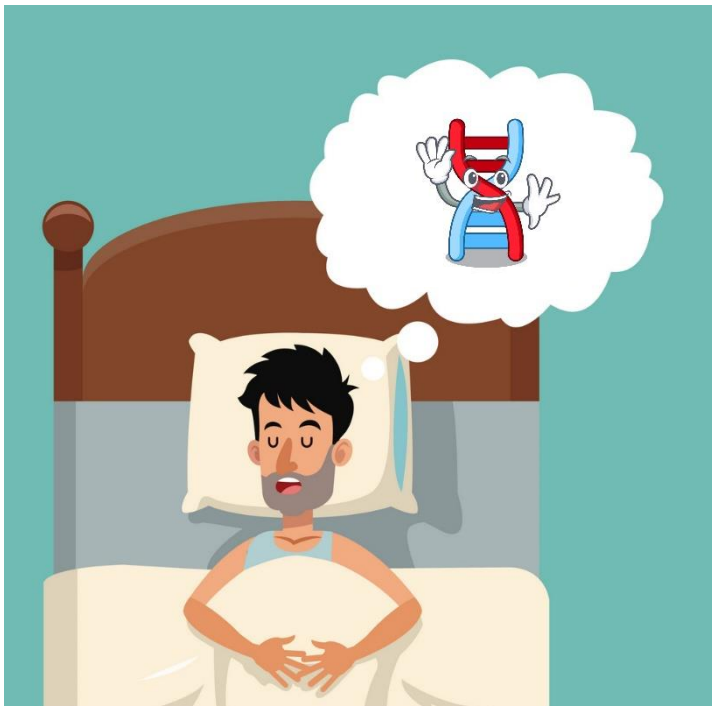
-G.Nikhila Chowdary,

Batch of ’21.

ARE DREAMS INHERITED?

Have you ever thought about why we dream, what we dream and how we dream? Does everyone dream? Are dreams connected to genetics and inheritance?

The average person dreams at least 4 times per night with each dream lasting up to 20 minutes. Usually, there are four stages in sleep: 1, 2, 3, and REM (rapid eye movement) sleep. REM sleep is the stage where most of the dreaming process happens and in this stage, the brain works as active as it does when a person is awake, but the body is paralyzed.



Recently, an international research team at the RIKEN Centre in Japan has identified a pair of genes that regulates how much REM and non-REM sleep an animal experiences. Two genes Chrm1 and Chrm3 have been discovered in mice associated with dreams, according to a study published in Cell Reports. It was found that the mice were incapable of experiencing REM sleep when these genes were removed. Even though there is no way of assessing whether the mice are dreaming or not, it can be said if they are in REM sleep or not. The assumption is that if the mice don't have REM sleep, they can't dream.

The knockout of Chrm1 fragmented REM sleep and the knocking out of Chrm3 reduced the length of non-REM sleep. When both genes were knocked out, the mice failed to experience REM sleep, but survived nonetheless. A study deprived rats of REM sleep and found that they lived only for five weeks as a result. The study illustrated the importance of REM to our overall brain health. The problems with REM sleep are linked to Dementia, Parkinson's disease and other neurological disorders. Till now they didn’t find any person who has Charcot- Wilbrand syndrome (loss of ability to dream). But a 73-year-old lady temporarily got affected by Charcot- Wilbrand syndrome when she got a stroke which stopped few of her brain functions.

Possible answers about what we dream include:

- Representing unconscious wishes.
- The things which affects us more in a day.

But there is also a case where inheritance plays a crucial role in what we dream about. This seems to be true when we dream about something which we are not aware of or in case if we find any similarity of dreams among closely related people.

In a research, scientists found that identical twins even experience nightmares on almost the same frequency. In a study which involved nearly 4200 non-identical twins and **2700 identical twins**, they found that identical twins are twice as likely to have the trait of having frequent nightmares as fraternal twins, because they are formed from the same zygote.

Not only in case of twins, but there are some situations where parent and offspring's dream the same. Does it make sense? Yes, of course by genetic memory.

Genetic memory is a memory present at birth that exists in the absence of sensory experience and is incorporated into the genome over long spans of time. A 2013 study from Emory University found that mice which are trained to fear a specific odour, pass their emotions on to their offspring's and future generations.

Fear isn't the only thing that gets imprinted in our genes. Experiments showed that a traumatic event could affect the DNA in sperm or egg and alter the genes, brain and behaviour of subsequent generations. The findings provide evidences of "transgenerational epigenetic inheritance". So, DNA is perhaps just one form in which some memory of parents' lives are stored, re-expressed in dreams and hypnotic regression.

At the same time, key emotions and memory-related structures of the brain are reactivated during REM sleep as we dream.

According to an article published on Reddit, a writer who was 16 had no idea about her dad witnessing an accident when he was in his 20's where he saw a guy completely missing his head. This incident was strongly recorded in his memory and for years afterwards, he had terrible nightmares about the headless guy. Even when the writer was a kid, her father continued to have these dreams. But the writer says that without having knowledge about the accident she used to get dreams on people ripping off their heads.

From this case, we can analyse genetic memory which had been passed to her from her father. As they are present in her memory, she can dream about it. In this way, our dreams sometimes might be inherited.

- P.Jaya Vasavi,

Batch of '21.

HIERARCHICAL DESIGN OF ARTIFICIAL PROTEINS AND COMPLEXES TOWARD SYNTHETIC STRUCTURAL BIOLOGY

All living organisms are made up of bio molecules like proteins, nucleic acids, sugars and lipids. Proteins are important components and are made up of amino acids. Synthetic biology is an interdisciplinary field that aims to re-design and fabricate biological components. Structural biology includes molecular structures and dynamics of biological macromolecules like nucleic acids and proteins. The integration of both, synthetic biology and structural biology is important for the structure, function and action of bio- nano machines. The DNA comprises of four bases A,T,G and C. The genetic content encoded by the gene is transcribed into RNA by the single replacement of T to U. RNA is then translated into a protein which is made up of 20 different amino acids. Approach of hierarchical design of artificial proteins and complexes can be achieved by using different methods.

Hierarchical design of tertiary structures of artificial proteins:

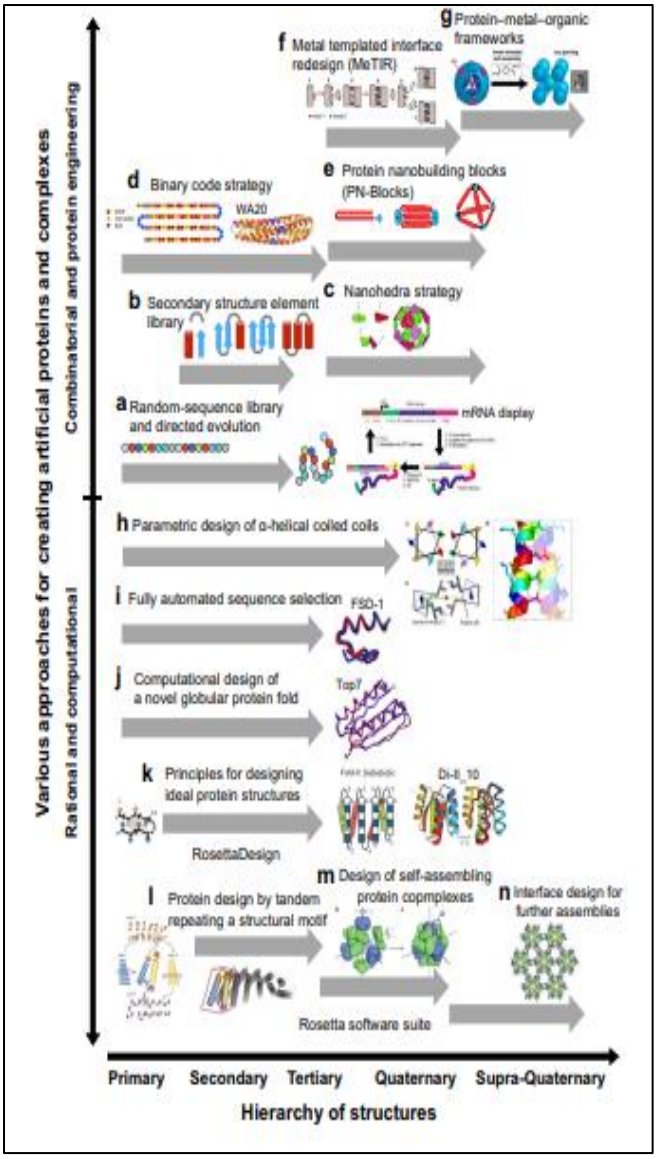
Primary, secondary, tertiary, quaternary are four different levels of protein structures. Primary structures are determined from the sequences and are considered as libraries for tertiary structures which are folded structures of the *de novo* proteins which use secondary structures like alpha helix and beta sheets with binary patterns of polar and non polar residues. This technique is called binary strategy. Protein domains have evolved from pre- existing polypeptide modular segments, secondary structure elements and supersecondary motifs derived from exon shuffling.

Combinatorial and directed evolution approaches for creating functional artificial proteins:

They do not require atomic-level structural information. The 3D protein should have good expression, solubility and assessment of molecular functions. Protein libraries should be selected based on the stably folded proteins. The genetic and phenotypic information have a link between them because of which the library of genes can be translated into corresponding libraries of gene products of functional similarity. The identification of the target binding site for a protein can be done by the binding activity and their encoding DNA sequences are captured using an immobilized target. In this way, non-binding libraries can be neglected.

Artificial binding proteins generated by protein engineering and combinatorial approaches:

Specific binding proteins are essential for diagnostic and therapeutic applications. There are many artificial proteins which



are created and they bind to a specific site.

They are highly repetitive structures of repeat proteins and repeated structural motifs are involved in protein-protein interactions. Macro surfaces are created by the variation of specific positions. By using different types of repeats such as Armadillo repeats, HEAT repeats and pentatricopeptide repeats an artificial repeat library is created. An artificial peptide library by the cyclized helix-loop-helix as a molecular scaffold for high through put screening.

Hierarchical design of quaternary and supra structures of artificial protein complexes:

These approaches include protein complexes and nano structures constructed from self-assembling fusion proteins, 3D domain-swapped oligomers and metal directed self-assembling proteins. Protein interactions can be achieved by modifying inter molecular helix-helix interface residues. By using symmetry, large molecules can be built up to icosahedrons and by using nanohedral Symmetry various materials can be created. In most cases protein-protein interfaces are stabilized by the metal ions in quaternary and supra molecules. It can be described by the metal-directed protein self-assembly and metal templated interface redesign.

These are some approaches to design and construct artificial proteins and protein complexes. These protein complexes facilitate the function of nano bio-materials. There is a structural and functional relationship in proteins. The structural aspects of protein are used to design specific binding proteins for therapeutic purposes. The production of protein from DNA has been related to high throughputs, designing of host cells and NGS.

-K. Geetha,

Batch of ’20.

ROLE OF BIOINFORMATICS IN CONTROLLING GREEN HOUSE EFFECT

The atmosphere is made up of a thick layer of gases, which makes the earth suitable for organisms to live. These green house gases (carbon dioxide, nitrous oxide and methane) are getting thicker, which warms up the atmosphere causing climate change. Recent studies show that microbes play a major role in biogeochemical process. The most abundant photosynthetic microbes are *Prochlorococcus* and *Synechococcus* and around 100 million microbes are present per litre of sea water. They remove about 10 billion tons of carbon and thus help in reducing the level of green house gases and climate change.

Here comes the role of bioinformatics which helps to analyse the genome of microbes which are useful in several areas of energy production, environmental remediation, toxic waste reduction, and industrial processing by using various genome sequencing and metagenomic projects, leading to generation of large amount generic data. The project results shows that six microbes live under extreme temperature and pressure. One of those is *Prochlorococcus* with the genome size of 2 Mb and lacks *phycobilisomes* that contains chlorophyll b as its major accessory pigment, responsible for a significant fraction of photosynthesis in the world's oceans and plays an important role in carbon cycle whereas *Synechococcus* contains chlorophyll a and phycobilins that are more typical of *Cyanobacteria*. The use of microarray technology for the analysis of gene expression patterns will give us knowledge about how these microbes manage the environment of the oligotrophic oceans. The study of genomes of these microbial organisms, which is possible through bioinformatics, helps in proposing ways to decrease the carbon dioxide content.

-Sweedha T

Batch of ’20.

CHILD MISHANDLE MIGHT ALTER DNA!

When you are in pain, a moment is an eternity and there was no one to tell and nowhere to hide keep the pain to myself, while a part of me died, this is what the victims think about thyself when there are abused .Now-a-days if we turn on the news there are reports of child abuse every minute. Studies from Harvard interestingly identified child abuse might alter the DNA of the victim and also the biological system. Epigenetic marks- which operates the genes functionally to on/off and differences have been found on the DNA methylation-DNAm (blood,saliva,brain tissues) as a mechanism by which child abuse increases the risk of neuropsychiatric and cardio metabolic diseases. If the DNAm gametes are associated with fertility, they affect the development of the future offspring. Interestingly, animal models are also examined by researches in which the effect of the sperm DNAm can change nutritional status, endocrine disrupting hormones, other pollutants and major psychological trauma which can pass on to the offspring through gametes. There were no prior studies related to the psychological stresses on DNAm, but it was recently identified that the psychological stress on the DNAm in humans lower the quality of semen, motile sperm concentration and head displacement and also identified for the prepubescent period become sensitive to environmental influence. At present experiments were developed where the change in sperm DNAm is calculated as the non-clinical longitudinal group of men in which the set of observations correlated by using the principle component (PCs) able to the opposed regions sites of the DNAm.

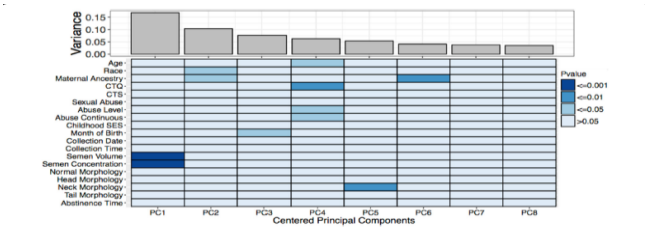


Fig. 1 Principal component 4 (PC4) was associated with childhood abuse exposure (one sample per participant, $N = 34$). PC4, representing 6.24% of the variance present in the methylation data, was significantly correlated ($p < 0.05$) with childhood abuse exposure. Darker regions signify stronger correlations between variables and principal components (N probes = 439,746). Normal sperm morphology is characterized, beginning at the head and moving toward the tail. Thus, "head morphology" is the % of sperm in a sample with normal heads, "neck morphology" is the % with normal heads and necks, and "tail morphology" is % with normal head, neck, and tail. Abstinence time is the time between the sperm donation and the most recent preceding ejaculation. PC principal component, CTQ Childhood Trauma Questionnaire, CTS Conflict Tactic Scale

The researchers took samples from victims at different ages and time period and performed the task accordingly. The experience of the physical, sexual, emotional abuse were measured as four items (hit, insulted, yelled at, punished cruelly) as childhood trauma questionnaire- CTQ (individual had an event that meets the criterion A) and responses of CTQ were summed and then divided into quartiles based on the distribution of the group. Physical and emotional abuse can be measured by conflict tactics scale(CTS). The responses for the CTS and the CTQ ranged from never to very often.CTS were skewed therefore divided as 2: highest-25%(HIGHEST LEVELS OF CTS OR CTQ), 0:lowest-50%(NO ABUSE) and 1:next-25%(MEDIUM ABUSE). They examined the covariates (collection of (time, date and age), percent normal morphology and hypothesized mediators as smoking, trauma, mental assault in which association of child abuse alter the DNAm which keeps the victim health at risk. The analysis is made by principle component analysis, machine learning analysis, Dmrs analysis, pyrosequencing methylation conformation, exploratory analysis and by this analysis results are displayed accordingly.

Table 2 Differentially methylated regions (DMRs) associated with childhood abuse exposure

Cluster name	Number of significant probes	*p*-value	FDR	Average $\Delta\beta$	Max $\Delta\beta$
ARL17A	3	1.54E-10	2.43E-07	−0.29	−0.35
MAPT	8	7.66E-10	7.99E-07	0.132	0.173
CLU	11	9.82E-05	1.04E-02	0.08	0.139
LRRK1	3	1.03E-17	1.19E-13	0.103	0.12
PRDM16	7	4.13E-05	6.95E-03	0.094	0.148
TCERG1L	3	1.60E-04	2.26E-02	0.131	0.147
CFAP46	5	2.09E-04	2.61E-02	−0.108	−0.122
MIR5093	4	2.52E-07	1.49E-04	0.108	0.128
TAF1B	3	6.47E-05	1.19E-02	0.148	0.194
DLL1	5	4.13E-05	8.52E-03	0.115	0.135
SYCE1	3	1.14E-09	1.31E-06	0.083	0.114
NDFUA10	3	1.60E-06	6.80E-04	0.119	0.138
SDK1	8	1.60E-04	1.93E-02	−0.091	−0.12

Statistically significant DMRs were discovered using DMRcate (FDR ≤ 0.05), had a mean $\Delta\beta \geq 5\%$, and were verified using replicates. *P*-value, FDR, and mean $\Delta\beta$ for each DMR are the mean across all probes within the DMR. $\Delta\beta$ values were calculated as the difference between the mean β for high and no childhood abuse.

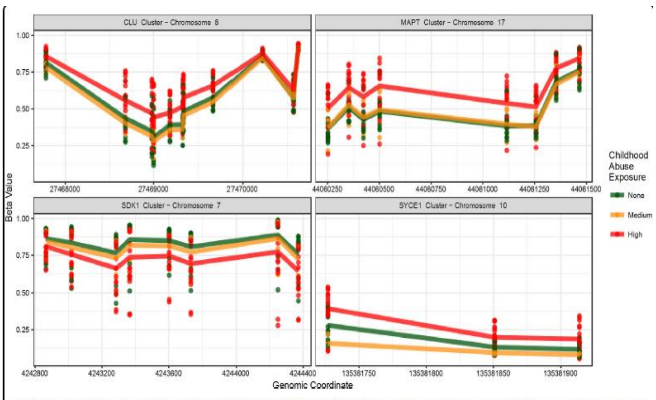


Fig. 2 Four genomic regions differentially methylated by childhood abuse. Differentially methylated regions (DMRs) were defined as regions that differed statistically by abuse exposure at an FDR ≤ 0.05, had a mean $\Delta\beta \geq 5\%$ across probes, and were confirmed using replicates. The "*CLU* cluster" includes the 5′ UTR transcription start site and part of the gene body spanning 2.8 kb. The "*MAPT* cluster" is located in the gene body and spans 1.2 kb. The "*SDK1* cluster" is located in the gene body and spans 1.5 kb. The "*SYCE1* cluster" is located in the 5′ UTR and spans 200 bp

Reference: *Roberts et al. Translational Psychiatry (2018) 8:194*
DOI 10.1038/s41398-018-0252-1

-A Sai Bavitha,

Batch of ’20.

BASEBALL STATISTICS MEETS COMPUTATIONAL BIOLOGY

To become a computational biologist, we don't have to go to a museum, a planetarium or any lecture, knowing how to analyse a baseball game is enough. The game statistics gives an overview of how to think about computational biology by obsessively analyzing baseball statistics. Let us start with the history of how the data is collected.

So, baseball statistics are collected at an individual game where people would just keep score on a score sheet and DNA sequence were obtained by a cumbersome method where you take a piece of DNA and performed basic biochemistry techniques on it and ran it out on a sequencing gel.

The obtained DNA sequence was recorded by writing it down on a single piece of paper. Eventually the results would be disseminated in the case of baseball statistics through a box score in a newspaper. In the case of DNA sequence, they were printed in a journal article but it was not really until those statistics and the sequencing data became systematically organized in the mid-1970s after almost a century of collecting baseball statistics.

Nearly a decade later of collecting DNA sequence data, visionaries in both fields realized that it would be valuable to put this information altogether in one place where you could look up the past performances of any baseball player or you could look up the DNA sequence of any gene whose sequence had been determined. The databases were used to make some predictions about baseball players and how many home runs Dustin Pedroia might hit and for the DNA sequence to predict the function of different proteins.

We can take the statistical profile and search it against a database of known players against the previous players who are older than Pedroia and look for similar

statistical profiles of players up to the age of 27. For example, Jose Vidro hit 15 home runs when he was 28, Tony Lazzeri hit 15 home runs, Chuck Knoblauch only hit 9 and Robinson Cano hit 28. Therefore, it is a range from 9 to 28, but we can see that essentially the average. Second baseman with a performance like Dustin Pedroia's up to the age of 27 hit around 18 home runs. So we can make the prediction that Dustin Pedroia is going to hit 18 home runs in 2012.

We use a similar logic to ask questions about proteins, in this case, we have sequence of a protein, all of the amino acids that encode it and we want to ask, "What does this protein do? What does it look like? In addition, is there anything different about its function?" Therefore, we take the protein and look against a sequence database using computer program called BLAST and out of the millions of sequence that we have for proteins, we find the ones that match up the best to our protein of interest.

Align them against each other to look for region of the protein that has similar amino acids sequences for which we know the 3D structure. Moreover, it is a mutation that individuals carry which sometimes are highly prone to develop a serious disease like amyotrophic lateral sclerosis. The predictions follow a similar logic like to find similar players/protein. Build a model including Historical/evolutionary comparisons, features unique to players/protein and to infer performance/function. Every single major league pitch is actually tracked from the time it leaves the pitcher's hand until it either hit or crosses the plate and we try to learn something from the velocity of the pitch or pitcher's arm angle.

In case of a DNA sequence data, a new generation of DNA sequencing has increased our ability to sequence the DNA by five or six orders of magnitude. We are still trying to figure out exactly what we can learn by sequencing every piece of DNA that exists anywhere on the planet. Development of new tools is required by learning something from that data.

Thus, we expect that the same types of tools that are developed to study the PITCH f/x data in baseball are going to use the same types of underlying algorithms and logic as the tools that are developed to analyze new sequencing data.

-Nanda Kumar M R,

Batch of '19.

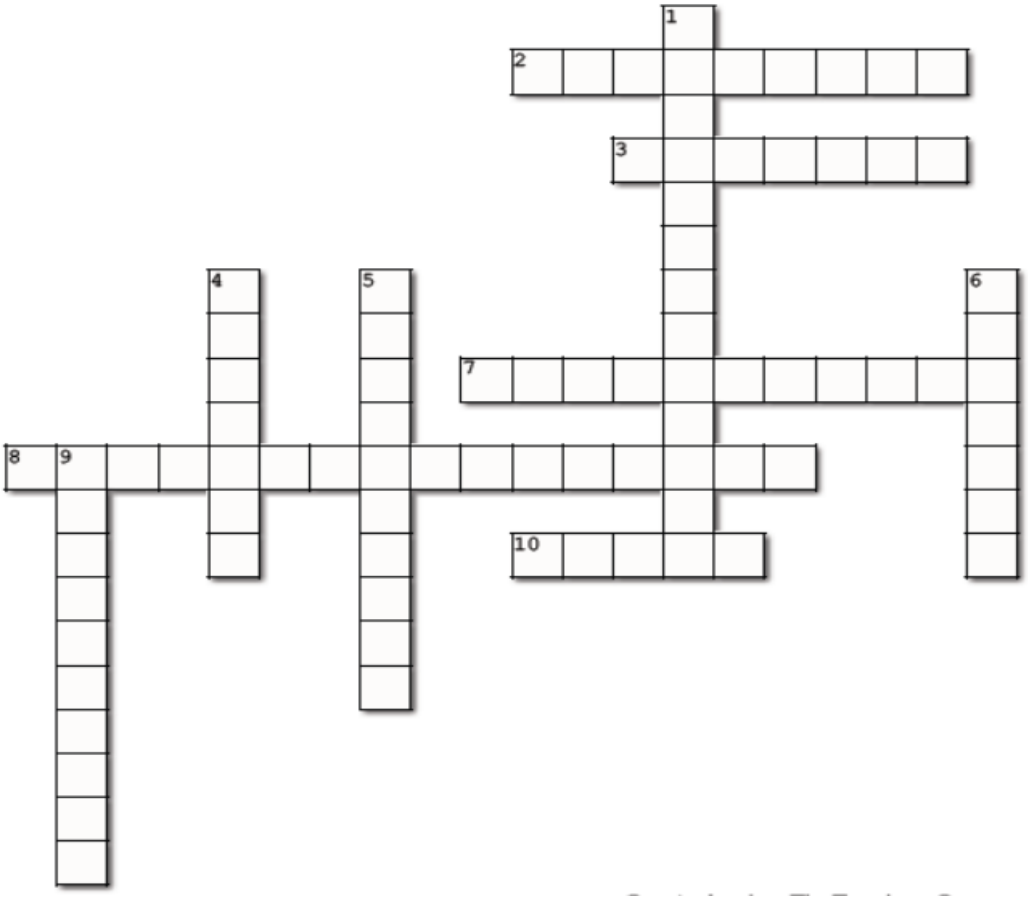
SKETCHES

- Kavya Karthic(Batch of '21)

-N Ayshwarya(Batch of'21)

-Kavya Karthic(Batch of '21)

CROSSWORDS



Across

2. I am a phenomenan of Speciation event in evolution

3. I am also a genetic material independent of chromosomal DNA. You can spot me in Bacteria easily

7. I can explain how to regulate gene expression without altering the gene sequence

8. I can tell you how a drug affetcs an organism

10. I can locally align Central Dogmal Sequences

Down

1. I can relate anyone in the history

4. We are a part of protein family and we all accelarate Chemical reactions

5. I am a multiplex lab-on-a-chip capable of doing assays with biological samples in HTS method

6. I am a hexamer protein at rest.I become a monomer protein when I am active and I am related to Diabetes

9. I am a tetrameric protein and I carry oxygen.

-Abirami R

Batch of ‘19