APS COVID Webinar: December 16, 2020

Immune interactions and SARS-CoV-2 Evolution

Benjamin Greenbaum, Memorial Sloan Kettering Cancer Center

- Viruses are obligate parasites
    - Depend on hosts to replicate; not alive without a host
    - Evolve much faster than other genomes due to poor proofreading
    - Have diverse sets of genomes and thus encounter different aspects of hosts machinery
- What immune interactions matter for virus evolution and pathogenesis?
    - Viruses interact with immune system both via molecules detected by the system and the creation of proteins that disrupt immune functions
    - A novel virus that has evolved in a different host can interact with the human immune system in a very different way
- **Theories of pathogenesis in COVID19**: (neither proven yet, but encourage physicists to explore!)
    - **Dysregulated innate immune signaling**
    - **Dysfunctional T-cell response**
- SARS-CoV-2 basics:
    - 30kb virus (very long due to proofreading)
    - Smaller error rate than other RNA viruses (because of proofreading)
        - Normally ~1 mistake every 10,000 bases, this one is order of magnitude lower
        - Good news for vaccines! Less mutations
    - Has a somewhat broad host range (bats, pangolins, humans)
- Immune system in COVID patients is highly dysregulated
    - These dysfunctions are theorized to lead to the most severe cases
- Early phases of immune sensing are critical in coronaviruses pathogenicity
    - In severe cases, somehow the connections are messed up; i.e. poor signaling, over-signaling
- **Innate immune interactions**
    - What can be sensed in theory?
        - Structural features found in known pathogens
            - Nucleotide bias (sensor=RIG-I)
            - dinucleotide bias (sensor=ZAP?)
            - dsRNA (sensor=MDA5)
    - Why do we think this has to do with pathogenicity?
        - Comes from study of 1918 flu and why it was so bad
            - Impression from data suggested an unusually high fatality rate in young adults
                - Implied something particular about how the virus played out in humans
            - Had a high type-I IFN and pro-inflammatory cytokine production
            - Similar idea holds up in bird flu
            - Influenza virus: RNA- virus, evolved via drift and shift, no proofreading, segmented into 8 different segments (SARS-CoV-2 is one long strand), much wider host range than SARS-CoV-2, more avian reservoir than other mammal
            - Paths of 1918 HA(H1) from birds – was also kind of a pandemic in swine

        - Was very closely related to a flu virus living and evolving in humans for 90 years
        - Why? Flu evolves through reassortment
            - Most human pandemics have been associated with reassortment of a human and non-human strain
- Innate immunity is based on recognizing patterns in foreign viruses
    - Often acts on non-coding features
    - Pattern recognition theory: PAMPs are a signal to the host produce signaling molecules that act as an alert to the host of the foreign virus
- Karlin genome statistics:
    - Karlin would chop a genome from different hosts, and come up with a sequence to be able to sort the genomes based on their patterns/words
    - Get better resolution for more longer words
    - Can do this with amino acids
- Looked at dinucleotide evolution in influenza – found the non-self becoming more self
    - Also noted the 1918 strain seems to come from the middle of the avian flu evolutions
- Human RNA viruses 'mimic' human RNA
    - Viruses that evolve in humans seem to be pressured to respect the usual human nucleotide patterns
- Binned H1N1 sequence history and context
    - Used the context to predict which viruses might be immunogenic
- Selective and entropic forces
    - Views evolution of motifs as competition between selective and entropic forces, latter of which randomize constrained sequences
    - Iterate using transfer matrix method until convergence
    - Calculate the number of motifs until convergence
    - Formalism allows for evolutionary dynamics
    - Evolutionary forces can then select forces and relaxation
    - Can use this information to predict how long it will take a virus to become self
    - **Key prediction**: evolutionary distances -> receptor differences
        -> Potential 'over or under presentation of PAMPS in a pathogen
        -> Differences in pathogenicity
    - Can use this to compare properties of host and virus genomes
        - CpG pressure seems to be really important for driving evolution
        - Exciting development is people looking at the associated receptors
            - i.e. ZAP molecule is sensitive to virus and inhibits the CpG-enriched HIV-1 replication
            - i.e. differential specificity in avian ZAP – effect of ZAP on CpGs is different between humans and birds (chickens); tells us virus will experience greater selective pressure when it jumps to humans
- This type of theory and logic goes into a lot of RNA vaccine design
    - Moderna's mRNA vaccine approach closely mimics a native viral infection leading to B and T cell responses
    - Random piece of RNA introduced to a cell will sometimes stimulate immune response and sometime not because of this innate response

        - Similar concept of finding the sensor, blocking and modifying the RNA, and then delivering with more effectiveness
    - Does all this innate immune interaction theory matter for SARS-CoV-2??
        - SARS-CoV-2 has a very biased evolution
        - Viruses display an APOBEC like mutational bias
        - Prominent given the low mutation rate compared to HIV or flu (proofreading)
        - Looked at CpG/UpA distribution..
            - CpG content in SARS-CoV-2 was extremely low compared to other cornonaviruses
            - Sliding window of the virus shows a lot of heterogeneity in the CpG dinucleotides
                - End piece has high concentration of CpG dinucleotides
                - SARS-CoV-2 can control its gene expression by making shorter pieces of RNA which are then translated
                - So virus can translate the high concentration CpG dinucleotides very quickly as it replicates
            - Clear C->U bias and prediction of ZAP binding
                - Validated today! - There is a fitness advantage for SARS-CoV-2 in cells where ZAP is knocked out
- **T-cell interactions**
    - HIV is the canonical example:
        - ssRNA + retrovirus
        - approx. 10 kb genome
        - complex life cycle
        - infects CD4 (helper) T-cells (primary target of the virus)
        - comes from our nearest evolutionary neighbors (chimps)
        - similarly high error rate of transcription
        - classic picture is the collapse of T-cells and evolution into AIDS
        - appears more 'self' than influenza
    - Antigen presentation machinery
        - Proteins are presented through cellular machine on the edges of the cell where they can be recognized by T cells
    - Many of the severe COVID cases seem to be associated with T-cell dysfunction
    - There is enormous diversity in HLA molecules
        - There exist really great computational tools to estimate whether a peptide is capable of being presented to the immune system
        - Can work your way back to physical properties of the residues responsible
    - Sorted outcomes by MHC type
        - Found that there were particular HLA types in patients with more severe cases
        - Saw that the ability of this HLA type to present the peptides was much lower – makes it hard to activate a T cell response
- Really interesting active work in the lab is **manipulation of non-self to self**

QUESTION AND ANSWER:

- Can you speak more on the current work on RNA vaccines (and speed)?
    - There is an understanding that work on vaccines is pretty inefficient
    - RNA vaccine work actually started with a focus on vaccinating against cancer
- What is benefit of RNA vaccines over DNA vaccines?
    - DNA vaccines add another step to the process, though innate immunity step is likely easier to solve
    - DNA vaccines evolved first because there is less work in optimizing, but one big problem is injecting DNA which evokes the problem of genetic manipulation
- Since the RNA vaccine technology is going to be extremely important in future, what do you think physicists can bring to the table to advance this technology?
    - Because of the fact that initial use was in cancer, they've been around for a long time – lots of clinical trials for safety and effectiveness
        - Low effectiveness in cancer – reason is likely that the ability to predict the right antigens in cancer prevention is very difficult
        - This application to viruses is probably a better use for RNA vaccines, and will hopefully help develop cancer vaccines better as well
    - Engineering problem of delivering things more efficiently – very important problem!
- Would you expect that the efficacy of the vaccines will decrease with time?
    - This is a big open question
    - Helps that it's a slowly evolving virus
    - Particular mutations seem to cause the virus to transmit differently – might cause large changes
    - Question is whether the T cells which learned the virus are still going and whether new antigens will have been introduced
- On evolution of SARS-CoV-2, the model so far has been that coronavirus circulated in bats, probably went through pangolin, and then into humans. Many people argue that some of the data does not validate this theory. Particularly, when we look at how SARS-CoV-1 evolved much more in humans than SARS-CoV-2. Suggests a theory that there could have been a lab accident where the virus was already being evolved in the lab and someone caught it. Which theory do you think is correct? How can physicists help determine this?
    - Note that there doesn't seem to be any evidence of engineering of this virus
    - Hard to tell either way since we don't do great surveillance of reservoir hosts
    - Very hard to prove the sources even of historical viruses
    - Strong epidemiological evidence in many cases, but there's really no substitute for reservoir surveillance efforts
    - Physicists may have a lot to offer in terms of sorting through that surveillance data and analyzing the patterns