So it’s a couple minutes after the hour, so why don’t we get started?
Welcome to grand rounds.
We have two really interesting talks today.
For those of you who don’t know me, I'll just.
Point out that my name is Ryan Crop.
I’m very new to Yale as the new medical director for the Clinical Trials Office.
And it’s pleasure to start meeting everyone.
So we have two took off.
As I mentioned the 1st.
Benjamin Turk has been
kind enough to join us.

He’s an associate professor of pharmacology and director of Medical Studies and pharmacology. He’s a member of the CSN program.

His graduate work was in biological chemistry at MIT and then a postdoc with Luke Cantley also in Boston,

and he works understanding molecular mechanisms underlying signaling pathways and how they’re organized into large networks,

Setting protein modifying enzymes and particularly kinase and proteases that are important in
00:01:15.445 --> 00:01:18.322 in signaling networks and today he
00:01:18.322 --> 00:01:20.866 will be talking about the aquaponic
00:01:20.866 --> 00:01:23.418 not kinase signaling network.
00:01:23.420 --> 00:01:24.428 Thank you Doctor Turk.
00:01:25.800 --> 00:01:29.848 OK, thank you. I will share my screen.
00:01:32.410 --> 00:01:33.300 See.
00:01:36.230 --> 00:01:39.398 OK can you see my screen in the pointer?
00:01:39.400 --> 00:01:41.314 Yeah perfect OK great.
00:01:41.314 --> 00:01:44.443 Well thank you for that introduction and
00:01:44.443 --> 00:01:47.517 also for the invitation to present our
00:01:47.517 --> 00:01:50.839 work on MAP kinase signaling networks.
00:01:50.840 --> 00:01:52.220 So as we all know,
00:01:52.220 --> 00:01:55.230 one of the hallmarks of cancer is
00:01:55.230 --> 00:01:56.983 uncontrolled cell proliferation and
00:01:56.983 --> 00:02:00.160 survival and cancer cells accomplish this,
at least in part through Co

opting signaling pathways that are normally activated downstream of peptide growth factor receptors.

I’m gonna be talking about one of the major arms of growth factor signaling the wrasse, RAF, MEK, Erk, signaling cascade.

So owing to high frequency mutations in rosanky genes as well as mutation, amplification of growth factor receptors, this pathway is amongst the most highly hyper activated in or more frequent most frequently hyperactivated in human cancers. And though the pathway has been the
subject of intense study for decades now, there are still some. Open questions in the field that our lab and of course many others are trying to understand and to sum up some of these questions that I’m really talking about today. One question, what are the functionally important components of MAP kinase signaling networks? So obviously the kinases that form the core cascade are have been well studied or well understood, but we have less understanding of other regulators of the pathway.
So for example the protein phosphatases.

That act on these kinases, and thus attenuate signaling through the pathway and it in addition.

We don’t have a complete catalogue of the substrates of Erk that act as the critical effectors in mediating the cancer cell phone. It type.

So in addition, one question we’re interested in is how specific connections are made between the kinases and the regulators and substrates in this pathway.

So there’s been a lot of really beautiful structural work emerging recently on the upstream components of the pathway,
NOTE Confidence: 0.92828127
00:03:47.810 --> 00:03:48.564 in particular,
NOTE Confidence: 0.92828127
00:03:48.564 --> 00:03:51.203 how Rask connects to RAF and MEK.
NOTE Confidence: 0.92828127
00:03:51.210 --> 00:03:52.040 But again,
NOTE Confidence: 0.92828127
00:03:52.040 --> 00:03:54.530 our understanding of the more downstream
NOTE Confidence: 0.92828127
00:03:54.530 --> 00:03:57.042 components where where we have these
NOTE Confidence: 0.92828127
00:03:57.042 --> 00:03:59.057 critical effector kinases and their
NOTE Confidence: 0.92828127
00:03:59.057 --> 00:04:01.179 substrates is is less well understood.
NOTE Confidence: 0.92828127
00:04:01.180 --> 00:04:02.311 And then lastly,
NOTE Confidence: 0.92828127
00:04:02.311 --> 00:04:04.950 one thing we know is that the
NOTE Confidence: 0.92828127
00:04:05.033 --> 00:04:07.493 persistent high level of activation
NOTE Confidence: 0.92828127
00:04:07.493 --> 00:04:11.231 of this pathway that one gets with
NOTE Confidence: 0.92828127
00:04:11.231 --> 00:04:14.221 new genic activation really doesn’t
NOTE Confidence: 0.92828127
00:04:14.221 --> 00:04:16.382 faithfully recapitulate the sort
NOTE Confidence: 0.92828127
00:04:16.382 --> 00:04:18.192 of normal dynamics of activation
NOTE Confidence: 0.92828127
00:04:18.192 --> 00:04:20.724 when we’d see in response to a
NOTE Confidence: 0.92828127
growth factor in a normal cell. And this can lead to a phenomenon that someone sometimes called network rewiring and how. New or which new connections are made in these networks and which connections are broken is something that’s important to know in terms of having a complete understanding of tumor cell biology. So I’m gonna tell two stories briefly today. The first has to do with the oncogenic map kinase signaling in Melanoma.
malignant melanomas are really driven by hyperactive Erk, and so about half of melanomas Harbor mutations in the BRAF gene. Most frequently the V 600 year Leal that leads to high level constitutive activation. And the remaining tumors have mutations. Most of them have mutations either in the NRAS, GTP ace, the NF One Ras. GTP is activating protein that negatively regulates the pathway or gain of function mutations in MEK, one which is just around stream of UVB and of course the dependence of melanomas on this pathway has really driven
the development and eventual FDA approval.
Of kinase inhibitors that target both B-RAF and MEK that are currently used to treat Melanoma, and while there is a high response rate for tumors that harbor B-RAF mutations, the problem with these drugs and really all targeted therapies is that the responses are not durable and patients will relapse within a few months to a couple of years. And the most common way that one sees resistance to these inhibitors is through reactivation of the Erk MAP kinase pathway. Despite the continued presence of inhibitor but one can also see
activation of bypass pathways like the PI3 kinase mTOR pathway leading to resistance and obviously there’s been a lot of interest in understanding these mechanisms of tumor cell resistance to these therapeutic agents. With the idea that if you understand how cells become resistant you might be able to devise new therapeutic strategies that might be more durable. So we got into this area through a genetic loss of function screen and SH RNA screen that one of my graduate students,
Eunice Cho conducted to identify genes that modulate sensitivity to MEK inhibitors in Melanoma cells and this work was published last year. People are interested in getting more of the details before I talk about the specifics of this research, I have to briefly plug the Yale Cancer Center Functional Genomics core that I Co direct with David Calderwood. And really, the the the the mission of this core is to facilitate these loss of function genetic screens. Core is to facilitate these loss of function genetic screens. CRISPR CAS 9 screens or SH RNA screens like I’m going to talk about and so
00:07:25.875 --> 00:07:27.716 hopefully this talk will give you a
00:07:27.716 --> 00:07:29.216 flavor of the kinds of information
00:07:29.216 --> 00:07:31.848 you can get out of these screens
00:07:31.848 --> 00:07:34.085 and inspire you to contact us and
00:07:34.085 --> 00:07:35.495 and set up your own.
00:07:35.500 --> 00:07:37.300 So I don’t have a lot of time to talk
00:07:37.300 --> 00:07:39.284 about the details of how the screen works.
00:07:39.290 --> 00:07:42.390 Needless to say, we.
00:07:42.390 --> 00:07:44.605 In introduce Melanoma cell line
00:07:44.605 --> 00:07:47.590 with a pooled SH RNA library.
00:07:47.590 --> 00:07:49.288 In this one.
00:07:49.290 --> 00:07:50.662 one targeting kinases and
00:07:50.662 --> 00:07:52.377 phosphatases and then we propagate
00:07:52.377 --> 00:07:54.470 in either the presence or absence.
One of two MEK inhibitors tromette never sell you met in IB and then we look at which hairpins become enriched or depleted from the population at the end of the screen and what this will should tell us our what jeans. Impact the sensitivity mekan hitters and hopefully identify additional genetic modifiers of of map kinase signaling. Assyrian threonine phosphatase called PPP six seed, and what you can see here is that amongst all of the hair pins that were in our library,
those that target PPP succeed were specifically enriched in the presence of either Solomon if or trim it nib. But under control conditions they were not enriched in the population. And what that means is that when you treat cells with a MEK inhibitor, they grow better if you knock down PPP 6C OK. So seeing PPP 6C as a hit in the screen really caught our eye and the reason for that is that something like 7 to 9% of melanomas have been shown through genomic analysis. Whole exome sequencing to harbor
what are thought to be loss of function mutations in PP6C.

So we thought identifying this phosphatase in the screen for modulators of drug sensitivity in Melanoma cell lines was probably not a coincidence.

So first thing we did was to try to verify this result.

So we derived PDP60 knockout cells through CRISPR CAS 9 gene editing, and sure enough, if we titrate in MEK inhibitor in this case a trim it and if you can see that knocking out PPP6C leads to substantial resistance to the inhibitor and the other thing that we’re observing,
which is kind of interesting,
is that actually sells deleted for
PPP6C grow more poorly than wild type
cell line than the wild type cell line.
But that growth is at growth effect is
actually rescued by low concentrations
of the MEK inhibitor and this is
reminiscent of a phenomenon that’s
been seen in in preclinical models,
that’s called inhibitor addiction and
basically what what this means is that
it’s it’s typically characterized
by cells having hyperactive map
kinase signaling and hyperactive
map kinase signaling is toxic to
cells and they can be brought back.

Down into the range that’s optimal for cell growth with low concentrations of an inhibitor,

and so that was it in a sort of an immediate clue that of what might be going on here.

That if loss of PPP6C caused hyper activation of MAP kinase signaling,

that would explain why you get resistance because it requires higher concentrations of drug to suppress the pathway enough to inhibit cell growth.

And also explain this drug addiction phenotype.

And sure enough, that’s what we see.
So basically, if we look at a number of distinct PPP 60 knockout clones, we can see profound hyperphosphorylation hyperactivation of MEK and of Erk and we can rescue that hyperactivation by expressing a wild type allele of PPP 6C but not a phosphatase dead allele that’s catalytically inactive. OK, and we extended these observations to a whole panel of cell lines, regardless of lineages we lookin cell
lines that either harbor BRAF mutations, or crass Oren RAST mutations with a couple of exceptions. We see that when we knock down PPP 60 by SH RNA, we get increased mech and or increased ORC phosphorylation. So we do think this is a general phenomenon, at least in the context of oncogenic map kinase. Signaling so. PPP succeed is a phosphatase and in experiments that I I I won’t have time to tell you about. We had ruled out activation of upstream components of the pathway and had a good handle on this PB6C acting at the level of MEK because it’s a
phosphatase may most straightforward explanation would be that it directly dephosphorylates Mac and we do think that’s what’s going on. So in vitro phosphatase assays we could show that.

Purified PP6P6C complexes. Candy phosphorylate MEK, but they don’t be phosphorylate Erk, so there seems to be some substrate specificity for the upstream component and probably more compelling we could detect at least an indirect physical interaction between Mac and PPP 6C. So PP6C is the catalytic
00:12:43.732 --> 00:12:46.009 subunit of holoenzyme that is,
NOTE Confidence: 0.681916948571429
00:12:46.010 --> 00:12:46.515 heterotrimeric,
NOTE Confidence: 0.681916948571429
00:12:46.515 --> 00:12:48.535 that includes regulatory subunits
NOTE Confidence: 0.681916948571429
00:12:48.535 --> 00:12:51.644 that have ascribed roles and binding
NOTE Confidence: 0.681916948571429
00:12:51.644 --> 00:12:53.716 to substrates and recruiting
NOTE Confidence: 0.681916948571429
00:12:53.716 --> 00:12:55.270 them for dephosphorylation.
NOTE Confidence: 0.681916948571429
00:12:55.270 --> 00:13:01.741 And we could see in komuna precipitation
NOTE Confidence: 0.681916948571429
00:13:01.741 --> 00:13:03.705 assays that pulling down any of
NOTE Confidence: 0.681916948571429
00:13:03.710 --> 00:13:06.022 the three regulatory subunits.
NOTE Confidence: 0.681916948571429
00:13:06.022 --> 00:13:08.189 I will bring down Mac but not so
NOTE Confidence: 0.681916948571429
00:13:08.190 --> 00:13:10.780 much with the catalytic subunit,
NOTE Confidence: 0.681916948571429
00:13:10.780 --> 00:13:12.850 sort of confirming a role for these
NOTE Confidence: 0.681916948571429
00:13:12.850 --> 00:13:16.190 regulatory subunits in in recruiting.
NOTE Confidence: 0.681916948571429
00:13:16.190 --> 00:13:18.528 MEC two to the complex.
NOTE Confidence: 0.681916948571429
00:13:18.528 --> 00:13:20.751 So I mentioned that PPP 6C is
NOTE Confidence: 0.681916948571429
00:13:20.752 --> 00:13:22.958 recurrently mutated in melanomas
and so we wanted to look at whether these mutations affected signaling through the MAP kinase pathway. And so we perform rescue experiments where we re-expressed series of the most frequently observed mutants in our PP60 knockout cells and what we observed is with a single exception that these mutants were either entirely or partially defective. In their ability to mediate mech dephosphorylation, so we conclude that these are likely partial loss of function mutations and it sort of makes
sense that they’re functioning to increase signaling through the core pathway that drives melanomas. That is, the MAP kinase signaling pathway. So, unfortunately, PPP 60 mutations are rare enough that we really don’t know the Clinical relevance of these mutations to pathway activation, but we were able to mine some data from C bio portal and it did appear as if there was a significant correlation between the mRNA expression level of PPP6C and the level of either phospho Erk or phospho MEK as seen in reverse phase protein arrays.
So we do believe that PPP 6C is modulating flux through the pathway in tumors and may be a factor that influences therapeutic response. OK, so in conclusion of this first part we’ve identified PPP 6C as a new player in restraining oncogenic map kinase signaling through dephosphorylation of MEK and that loss of function. Mutations of PPP 60 lead to hyper activated Erk signaling some of the open questions that we’re trying to pursue. Now, how is PPP 6C regulated?
So this phenomenon where PPP6C is required to restrain MEK activation has something that we really only see in the setting of oncogenic activation of the pathway. And that suggests to us that maybe there’s a negative feedback loop where pathway activation leads to activation of PPP6C towards the phosphorylation of MEK, and we’d like to understand how that happens. And of course, it may be that there are other signaling outputs substrates other than mech that are functionally important for tumors where you see lots of people pay 60 and we’re interested.
in trying to identify those as well.

So for the remaining time, I'm going to switch gears a little bit and move downstream in the pathway to do the kinase in the bottom of the map kinase cascade IIRC, and here we're going to be talking a little bit more about the structural basis for how connections in the pathway are made, and also some of these network rewiring phenomena they introduced at the beginning and so the work I'm going to talk about is the work of really talented graduate student who's.
Currently in the lab Julissa Torres
Robles and what she was interested in looking at our oncogenic mutations in Erk 2 itself or encoded by the map K1G.
So as I said at the outset, you have high frequency mutations in multiple cancer types of Rasen draft but at lower frequency you do see mutations in some of the downstream components. The Erk mutations in particular are sort of interesting because you don’t see them in the same tumor types that you do the Rasen draft mutation. So where, where as Rasen rap mutations you see in melanomas, colorectal cancers, lung cancers,
pancreatic cancer, the Erk.

2 mutations are largely restricted to squamous cell carcinomas, so about 8% of cervical squamous cell carcinomas have recurrent or two mutations and about 2% of head and neck. Squamous cell carcinomas have these mutations and they've attracted some attention in that setting because of potential association between the presence of those mutations.

And clinical responses to EGF receptor inhibitors.

So one of the things that kind
00:17:28.732 --> 00:17:30.820 of attracted us to this is the nature of these mutations.
NOTE Confidence: 0.950039684210526
00:17:30.820 --> 00:17:32.320 They’re sort of unusual when you compare them to other activating mutations and protein kinases that you see in cancer.
NOTE Confidence: 0.950039684210526
00:17:35.041 --> 00:17:37.387 So unlike say, BRAF mutations or EGF receptor mutations, these mutations don’t intrinsically hyper activate the kinase and they map at least in three dimensional space to a really interesting region of the kinase catalytic.
NOTE Confidence: 0.950039684210526
00:17:39.680 --> 00:17:41.039 So unlike say, BRAF mutations or EGF receptor mutations, these mutations don’t intrinsically hyper activate the kinase and they map at least in three dimensional space to a really interesting region of the kinase catalytic.
NOTE Confidence: 0.950039684210526
00:17:41.039 --> 00:17:43.757 BRAF mutations or EGF receptor mutations, these mutations don’t intrinsically hyper activate the kinase and they map at least in three dimensional space to a really interesting region of the kinase catalytic.
NOTE Confidence: 0.950039684210526
00:17:43.760 --> 00:17:45.580 these mutations don’t intrinsically hyper activate the kinase and they map at least in three dimensional space to a really interesting region of the kinase catalytic.
NOTE Confidence: 0.950039684210526
00:17:45.580 --> 00:17:48.310 hyper activate the kinase and they map at least in three dimensional space to a really interesting region of the kinase catalytic.
NOTE Confidence: 0.950039684210526
00:17:48.383 --> 00:17:50.567 all map at least in three dimensional space to a really interesting region of the kinase catalytic.
NOTE Confidence: 0.950039684210526
00:17:50.567 --> 00:17:52.504 space to a really interesting region of the kinase catalytic.
NOTE Confidence: 0.950039684210526
00:17:52.504 --> 00:17:54.829 region of the kinase catalytic.
NOTE Confidence: 0.950039684210526
00:17:54.830 --> 00:17:56.496 So this is a region that falls outside of the catalytic cleft.
NOTE Confidence: 0.950039684210526
00:17:56.496 --> 00:17:58.110 outside of the catalytic cleft.
NOTE Confidence: 0.950039684210526
00:17:58.110 --> 00:18:00.567 That’s known as the common docking group,
NOTE Confidence: 0.950039684210526
00:18:00.570 --> 00:18:02.215 and it’s called that because
it serves as a hub for protein interactions with ERP two, so this docking groove binds to a number of substrates of Erk, but it also binds to icks regulators, so the Mach one and Mach 2 which are positive regulators that phosphorylate and turn on or combined at this site, and the dual specificity phosphatase that dephosphorylates find it this site as well. So this sort of presents a little bit of a conundrum because I just told you that this is a really functionally important
part of the molecule yet, and so you might expect that mutations at this site would be loss of function. But of course just logically it would seem that mutations in, IIRC, would seem that mutations in, IIRC, gain of function and and the reason why this is is that these mutations actually cause selective disruption of these protein-protein interactions. So for example, we know that these cancer-associated Earth mutants are still able to interact with MEK one and MEK two, and so they can be activated normally, but they no longer interact with
the dual specificity phosphatase. So incels, this leads to an imbalance between their activation and inactivation, and you accumulate the hyper phosphorylated active form of the kinase, but that’s not all there is to it because it turns out that at least one of the major signaling outputs of Earth that is the Chinese risk is also broken by these mutations, so these mutants don’t interact with risk and they don’t phosphorylate risk, and so that makes you raised a few questions in our mind. So first of all,
what is the scope of interactions with Erk that are selectively disrupted by her mutations? We simply don’t know this at this point and from a kind of structural or biochemical standpoint. Why are some interactions broken and some spared something that we also don’t understand? In order to address this question, Jay Lisa conducted a proteome wide screen to identify sequences that can interact with the Erk docking group, and again I don’t have time to explain this in detail.
What we did was mine the human proteome for short amino acid stretches of amino acid sequence. That sort of had sequence similarity to known interacting sequences like you would find in Mach one and Mach 2. And prepared a genetically encoded library of about 12,000 sequences. So these are short sequences, fragments of proteins that are 14 amino acids long. And then we use those in a pooled competitive yeast. Two hybrid screening format and the bottom line is that you
00:20:50.000 --> 00:20:52.587 know similar to sort of an SH RNA or
NOTE Confidence: 0.9346002125
00:20:52.587 --> 00:20:55.093 crisp screen if we have a successful
NOTE Confidence: 0.9346002125
00:20:55.093 --> 00:20:57.660 interaction between Erk and the interactor,
NOTE Confidence: 0.9346002125
00:20:57.660 --> 00:20:58.900 this will become enriched
NOTE Confidence: 0.9346002125
00:20:58.900 --> 00:21:00.140 in the population overtime,
NOTE Confidence: 0.9346002125
00:21:00.140 --> 00:21:02.380 and we can detect this by next
NOTE Confidence: 0.9346002125
00:21:02.380 --> 00:21:03.020 generation sequencing.
NOTE Confidence: 0.9346002125
00:21:03.020 --> 00:21:05.090 So when we do this screen with wild type.
NOTE Confidence: 0.9346002125
00:21:05.090 --> 00:21:08.730 Work we can see that on gratifyingly,
NOTE Confidence: 0.9346002125
00:21:08.730 --> 00:21:10.655 all of the known interactors
NOTE Confidence: 0.9346002125
00:21:10.655 --> 00:21:12.580 interacting fragments that were in
NOTE Confidence: 0.9346002125
00:21:12.649 --> 00:21:14.669 the library actually scores hits.
NOTE Confidence: 0.9346002125
00:21:14.670 --> 00:21:16.326 They become enriched,
NOTE Confidence: 0.9346002125
00:21:16.326 --> 00:21:17.430 and furthermore,
NOTE Confidence: 0.9346002125
00:21:17.430 --> 00:21:19.789 if we align all of these sequences,
NOTE Confidence: 0.9346002125
00:21:19.790 --> 00:21:21.866 we can see a sequence motif.
A signature sequence that emerges that seems to be a common feature of sequences that interact with Erk. So a cluster of proline residues and a couple of leucine residues close by, and this is interesting in its own right because it tells us something about how Erk recruits its interacting proteins, but what about the mutants? So J. Lisa conducted this same screen with the two most recurrent cancer associated mutations, D321 and E322K and what we saw was kind of what we expected, which is that most of the interactions are preserved about 2/3 of the time.
interactors that scored his hits for wild type or also interact with the mutants, but about a third of them interacted. Only with the wild type kinase, and furthermore, when we look at the sequences that interact only with wild type, we actually lose this sequence motif that’s characteristic of Erk binders in general. And actually there’s very little distinguishing feature here, save for the significant selection of a single arching residue in the sequence. So we were a little bit flummoxed by this at first,
but first we just wanted to do some basic validation.

I'm starting to run short on time, so I'm going to go through this briefly. Basically, we could confirm that a sensually all of the sequences, but if we made synthetic peptides corresponding to these sequences that scored as hits in the screen, we could see that where we expected we saw differential binding in vitro to wild type versus mutant alleles of Erk, one of them in particular peptide coming from the protein ISG 20 had
particularly high affinity for Erk and showed the biggest differential binding. Between wild type and mutant forms. So we decided to take a structural biology approach to understand what was going on here in terms of how this interacted with her and with a lot of help from Titus Boggins lab here in the pharmacology department, Jay Lisa was able to solve the X-ray cocrystal structure of wild type work too. In complex with this fragment of the ISG 20 protein and I'm just going to zoom in on the key feature at the region of ISG 20 that binds to IRK. That is close to the hot spot for
these mutations we see that the peptide forms a single turn of an alpha Helix and that is enforced. That motive interaction is enforced by a sequence motif that involves a hydrophobic isoleucine residue and then two arginine residues position close by that actually make direct polar contacts to the acidic residues that are mutated in in cancer. And sure enough, if we then go back. And look at our sequences. That bound most preferentially to wild type the the top 9 sequences in the original used to hybrid.
screening data all have this sequence

motif and we could further confirm

that this motif was important for

but not to mutant forms of work

through in vitro binding assays

that we did with synthetic peptides.

So basically,

if we mutate any

of these three residues.

We greatly reduce the binding

affinity with wild type IIRC,

but we have no effect on the

already weak binding affinity

with the mutant forms of FERC,

presumably because the damage had
00:24:51.400 --> 00:24:53.810 already been done by those mutants.
NOTE Confidence: 0.935739000909091
00:24:53.810 --> 00:24:55.950 So we think we have a good handle on why
NOTE Confidence: 0.935739000909091
00:24:56.012 --> 00:24:57.384 some sequences interact specifically
NOTE Confidence: 0.935739000909091
00:24:57.384 --> 00:24:59.880 with wild type work and are broken.
NOTE Confidence: 0.935739000909091
00:24:59.880 --> 00:25:02.134 The interactions are broken with the mutants,
NOTE Confidence: 0.935739000909091
00:25:02.140 --> 00:25:04.282 but we’re now trying to do is sort of
NOTE Confidence: 0.935739000909091
00:25:04.282 --> 00:25:06.035 understand a little bit more about how
NOTE Confidence: 0.935739000909091
00:25:06.035 --> 00:25:07.709 this relates to tumor cell biology,
NOTE Confidence: 0.935739000909091
00:25:07.710 --> 00:25:11.492 And So what we’ve been doing is looking
NOTE Confidence: 0.935739000909091
00:25:11.492 --> 00:25:13.543 at some of the full length proteins
NOTE Confidence: 0.935739000909091
00:25:13.543 --> 00:25:15.093 that corresponds to its corresponding
NOTE Confidence: 0.935739000909091
00:25:15.147 --> 00:25:16.870 hits from the screen, and one that
NOTE Confidence: 0.935739000909091
00:25:16.870 --> 00:25:18.310 in particular that caught our eye,
NOTE Confidence: 0.935739000909091
00:25:18.310 --> 00:25:21.705 is the row GTPS exchange factor def.
NOTE Confidence: 0.935739000909091
H1, which has been implicated in a positive feedback loop for the Erk signaling pathway.

It's a known substrate of work, and we can confirm that indietro, but also also confirm that these cancer mutated forms of Erk are unable to phosphorylate. FH1, at least in vitro, and we're now following up. On these studies in head and neck squamous cell carcinoma cell lines to see if we can verify this result and understand what this means for tumor cell biology.

So to sum up this part, we've identified that cancer associated
mutations that map to these common docking 
groove of Earth 2 disrupt a subset of 
interactions and specifically those 
involving a particular sequence motif. 
And what we’re trying to figure out now, 
of course, 
is if selective engagement of these 
substrates is important for the phenotypic 
consequences of work to mutation. 
So with that. 
I will stop and thank the people 
who did the work I mentioned, 
Eunice Cho,
who recently left the lab graduated 
last year who had done all the work on
PPP 6C and the work on Earth mutants was conducted by Julissa Torres Robles.

I also like to point out my collaborators, David Calderwood, who’s my partner in all the functional genomics stuff.

Tice Boggins lab who helped us with the crystallography and Mark Gerstein lab that helped us with the.

Library design and computational analysis. And with that I’m happy to take any questions if we have time.

Thank you that that was great and really nice work and and a good advertisement for the functional genomics core ‘cause it looks like some
really impressive data we have maybe two or three minutes for questions.
If you wouldn’t mind just putting him in the chat while people are doing that,
can I just ask you a quick question about the PP6C study?
Is it worth going back and trying to redo your.
Knock down screen in a background of the PPP mutant contacts to see if there’s other.
Targets that could restore sensitivity to the inhibitors.
Yeah, I do believe so.
And actually one of the things that
we have planned is such a screen.

So the screen that we did before was a focus SH RNA library and what we're gearing up to do is a genome wide CRISPR screen where we compare wildtype cells with the PPP 60 knockout cells in the presence or absence of the MEK inhibitor, and so we're hoping to get out genetic modifiers that affect the growth of the PPP 60 knockout cells and one of the hopes is that we'll identify potentially other signaling outputs of PP6C that are important for growth and
NOTE Confidence: 0.900493566666667
00:28:29.218 --> 00:28:30.450 maybe drug sensitivity as well.
NOTE Confidence: 0.776560158
00:28:31.960 --> 00:28:34.600 You know, it seems like it makes that
NOTE Confidence: 0.776560158
00:28:34.600 --> 00:28:37.990 make sense in just one other question.
NOTE Confidence: 0.776560158
00:28:37.990 --> 00:28:39.362 I got a little,
NOTE Confidence: 0.776560158
00:28:39.362 --> 00:28:41.770 maybe I misunderstood in terms of the.
NOTE Confidence: 0.776560158
00:28:41.770 --> 00:28:44.070 The prevalence of these mutations
NOTE Confidence: 0.776560158
00:28:44.070 --> 00:28:48.029 in the in the in that phosphatase,
NOTE Confidence: 0.776560158
00:28:48.030 --> 00:28:50.270 and they I I thought you had said
NOTE Confidence: 0.776560158
00:28:50.270 --> 00:28:52.400 that they were relatively common.
NOTE Confidence: 0.624400154
00:28:54.030 --> 00:28:56.834 It’s it’s 7 to 9% depending on
NOTE Confidence: 0.624400154
00:28:56.834 --> 00:28:59.244 the study, so they’re they’re.
NOTE Confidence: 0.624400154
00:28:59.250 --> 00:29:00.636 They’re not as common it it’s.
NOTE Confidence: 0.624400154
00:29:00.640 --> 00:29:01.108 It’s actually interesting
NOTE Confidence: 0.624400154
00:29:01.108 --> 00:29:02.044 if you look at the data,
NOTE Confidence: 0.624400154
00:29:02.050 --> 00:29:03.514 they’re sort of the I guess
NOTE Confidence: 0.624400154
the fifth most common,

you know after the big guys and Ranson.

If one and I think P 53

ey they're next

and do they get enriched? Have you do?

Are there any databases of MEK resistant MEK inhibitor resistant samples that

you can look to see whether it’s

enrichment for that mutation? Yeah,

that hasn’t really come out of those studies.

A lot of those studies have been.

Looking at sort of individual

patients and you know people

have made patients right.

Zena graphs and things

like that and and done.
You know whole exome saying there's no. I mean because they’re not particularly common that it really has not come up as a bonafide clinical resistance mechanism.

OK, alright thank. Thank you again. Really nice work so why don’t we move on to our next presenter? Is Doctor Grace Kang, who’s an assistant professor in Department of Psychiatry and a member of our cancer Prevention and Control research program. She did her graduate work in clinical psychology at Saint Johns and in postdoc.
in adolescent addictions in the Yale.

A school of Medicine's division of substance abuse.

Her current research interests include understanding, substance use, health disparities among youth, and the use of social media for tobacco marketing and novel tobacco use behaviors among youth.

Her title is leveraging social media analysis to inform tobacco prevention.

Dr. Kang thank you for joining us.

OK, can you hear me now?
Yep perfect OK great thanks.

And you could hear you could see my slice here, right?

Yeah, OK, awesome, thank you.

Well, thank you so much for having me here today.

We're gonna really switch gears and talk about social media and youth tobacco prevention, so I will give a brief outline of what we'll talk about today.

So I'll first given out overview of why we should care about East.

figure prevention in the context.

of tobacco prevention and and,
and then the importance of leveraging social media to understand Easter youth behaviors and promotion and and then talk about limitations on current methods to analyze social media and then introduce how advances in new computational methods could be used to overcome some of these limitations, to understand cigarette content and social media. So cigarette smoking is a leading cause of preventable cause of death, disease, disability and death in the United States and we also know
that smoking causes cancer's of a variety of charts in the body. However, cigarette is just one type of tobacco product in the market. There are other types of tobacco products such as cigars, smokeless tobacco, E cigarettes, just to name a few, that Berry in harm. And here what you see is this is a graph from CDC and this shows. Different tobacco products and use rates across the decade and what you see is overall this decrease in tobacco use right and but this dotted green line here is increasing E cigarette
use over the years since 2014, E cigarettes have been the most commonly used tobacco product use among youth and in 2020 more than 4.5 million of the US youth are are using E cigarettes. And so when you take E cigarettes into consideration, the overall tobacco use rates. Is increasing among US youth? So for those who are not that familiar with E cigarette, I’ll just provide an overview of what a E cigarette is. There are many different types of E cigarettes on the market. These devices are not regulated,
so there is a rapid innovation such
different product characteristics and E
cigarette devices have evolved overtime.
It first started out with Cigalikes,
And then evolve into second
generation on devices like vape pens,
Third generations are these mods which
vary in how they’re it could be really
customized in very different ways,
and it could also excel large
amounts of excelled aerosol,
and then there is this pod mods here
that sort of varies and how it looks.
The most notable device you may have heard of is Jewel. They recently got popular because they use nicotine salt instead of freebase. So freebase nicotine is manipulated so that it has more of the harshness or kick the smokers like. The nicotine salt is manipulated by lowering the pH level so that it's not as harsh and allows for higher levels of nicotine. The problem with using nicotine salt is that it's easier to debate, you know higher levels of nicotine could be included in this product and therefore.
could be a risk because of his high level of nicotine. So once Jewel started really hitting the market and getting really popular, this fifth generation of devices started entering the market and these are disposable pod devices. They’re meant to be single use sometimes with multiple packs. They’re small, they’re discrete, they look like jewel they contain. They also contain salt, so which has high levels of nicotine and it comes in multiple flavors. And there’s a widely and importantly,
they’re cheap, so you might see a lot of these products on. Come in in your gas stations and other store convenience stores. So how do you cigarettes work? You know, even though these cigarettes vary in how they look, so the anatomy is is the same. So it has a component that holds that you liquid. It has a heating element. Any of the power power source in the form of batteries and is a mouthpiece in which the user could use to inhale the aerosol from the of the vape and in some in some devices,
just inhaling could activate the device.

So what’s in E liquid is made up of nicotine flavorings.

The base is made up of proper link like coal and vegetable glycerin,

as well as other additives.

So in terms of nicotine,

that’s the main drug.

So it stimulates the,

stimulates the central nervous system.

It raises blood pressure,

respiration, heart heart rate,

and releases a feeling of pleasure.

And the E cigarette that comes in Freebase comes in zero
00:36:11.754 --> 00:36:13.879 to 36 milligrams per milliliter.
NOTE Confidence: 0.806130926666667
00:36:13.880 --> 00:36:15.905 The nicotine salt on their
NOTE Confidence: 0.806130926666667
00:36:15.905 --> 00:36:17.120 marketed as percentage.
NOTE Confidence: 0.806130926666667
00:36:17.120 --> 00:36:18.872 So so for example,
NOTE Confidence: 0.806130926666667
00:36:18.872 --> 00:36:21.500 Jewel come as come as 5%,
NOTE Confidence: 0.806130926666667
00:36:21.500 --> 00:36:23.515 which is equivalent to about
NOTE Confidence: 0.806130926666667
00:36:23.515 --> 00:36:25.127 59 milligrams per milliliter.
NOTE Confidence: 0.806130926666667
00:36:25.130 --> 00:36:27.160 And you know the the issue with
NOTE Confidence: 0.806130926666667
00:36:27.160 --> 00:36:28.999 labeling is also very important,
NOTE Confidence: 0.806130926666667
00:36:29.000 --> 00:36:31.496 because you know 5% of anything
NOTE Confidence: 0.806130926666667
00:36:31.496 --> 00:36:33.160 just sounds little right.
NOTE Confidence: 0.806130926666667
00:36:33.160 --> 00:36:35.162 But if you actually look at the
NOTE Confidence: 0.806130926666667
00:36:35.162 --> 00:36:36.659 milligram per milliliter is actually
NOTE Confidence: 0.806130926666667
00:36:36.659 --> 00:36:38.119 very high level of nicotine.
NOTE Confidence: 0.806130926666667
00:36:38.120 --> 00:36:39.554 And this is what makes the
NOTE Confidence: 0.806130926666667
00:36:39.554 --> 00:36:41.050 nicotine is what makes addictive.
There are zero level of eliquids and E cigarettes available. However, I should say that that’s not very common. These E cigarettes come in many different flavors. There’s more than 7000 flavors. You know it comes in the typical like menthol tobacco flavor, but what’s really popular or you know, fruit candy store that desert kind of flavors. And also there’s also a lot of names that does not allude to actual,
you know food,

but like obscure names like

you know Unicorn milk,

or you know vampire blood

or things like that.

That gets people’s attention.

It is made up of chemicals.

And the people in glycol,

vegetable glycerin and the

combination of the two is used.

The ratio of the two is to create

either more aerosol or less aerosol

either more aerosol or less aerosol

either more aerosol or less aerosol

is used to intensify flavors or or

is used to intensify flavors or or

a lower the intensity of flavors and

nicotine or other chemicals added

such as other water and other chemicals.
So in addition to you know nicotine flavor flavorings, PG, VG, and other chemicals E cigarette aerosol have known or are shown to have heavy metals volatile organic compounds, and fine and ultrafine particles that can be inhaled deeply into the lungs by both by users as well as bystanders. The long term effects of this vaping is currently unknown. Why should we care about E cigarettes? So nicotine use among youth increases.
the risk of lifelong tobacco addiction.

And it could also increase the risk for future addiction to other drugs as well.

This is this E sticker.

Use is considered an epidemic in the United States, so it’s NIH, including NCIS.

Research priority priority is to prevent thicker E cigarette use.

In fact SCI has RFA specifically focus on preventing E cigarette use.

Among youth and has a collaborative. A grant that’s interested in particularly interested in E cigarette preventing E cigarette use.

and then lastly they also have
invested considerable resources into developing smokefree.gov, which has resources to help youth to quit E cigarette use. So we're thinking of how to prevent E cigarette use. We've got to consider a lot of factors right, so there are social, environmental, cognitive, and genetic influences that play a role in youth tobacco use. But we also know is that tobacco promotion, marketing, advertising is causally related to youth tobacco use and this has been well established.
and has been talked about in surgeon general reports.

So I’m going to focus on social media because now with the advent of social media, tobacco promotion really faces a unique challenge because social media is fast, it’s cheap, you could reach a lot of people at a quick speed and it doesn’t have sufficient to. To control its content. So it might not be that surprising to you to hear that you know social media is popular among youth. 90% of youth have used social media, 75% have at least one active social media profile and 93% report visiting
00:40:03.954 --> 00:40:07.160 on social media site at least daily.

00:40:09.645 --> 00:40:11.763 When it comes to understanding how E cigarettes are promoted to youth is so important to understand how it’s promoted so pro E cigarette content.

00:40:19.100 --> 00:40:21.372 Is on social media through paid ads and influencers promoting the products and on post from a share by their peers and other people?

00:40:31.032 --> 00:40:33.492 And recent studies have or are finding that use of social media among youth is associated with E cigarette use?

00:40:35.378 --> 00:40:37.207 types of social media platforms in our in our group or I’m going to
00:40:39.320 --> 00:40:40.712 present research findings specific
NOTE Confidence: 0.785546605
00:40:40.712 --> 00:40:43.107 to YouTube and I’m and I’m sure
NOTE Confidence: 0.785546605
00:40:43.107 --> 00:40:45.256 all of you have used YouTube so
NOTE Confidence: 0.785546605
00:40:45.256 --> 00:40:46.409 you’re familiar with it.
NOTE Confidence: 0.785546605
00:40:46.410 --> 00:40:49.920 YouTube is free online streaming service.
NOTE Confidence: 0.785546605
00:40:49.920 --> 00:40:52.320 Is used by 1.9 billion users,
NOTE Confidence: 0.785546605
00:40:52.320 --> 00:40:54.496 which is a third of all Internet users
NOTE Confidence: 0.785546605
00:40:54.496 --> 00:40:57.034 and people spend about a billion hours a
NOTE Confidence: 0.785546605
00:40:57.034 --> 00:41:01.999 day watching watching online YouTube videos.
NOTE Confidence: 0.785546605
00:41:02.000 --> 00:41:04.952 So the the data on the right.
NOTE Confidence: 0.785546605
00:41:04.960 --> 00:41:06.380 The graph here shows this is data from 2018,
NOTE Confidence: 0.785546605
00:41:06.380 --> 00:41:09.220 so it’s a bit old,
NOTE Confidence: 0.785546605
00:41:09.220 --> 00:41:11.131 but it shows that among teens YouTube
NOTE Confidence: 0.785546605
00:41:11.131 --> 00:41:13.639 is still popular and actually
NOTE Confidence: 0.785546605
00:41:13.640 --> 00:41:16.128 I think this year last year that showed
that YouTube is still popular among youth despite newer platforms entering. That’s popular among youth. We could also see that among those people who use they’re using YouTube often. Researchers have examined E cigarette content on YouTube to inform prevention. They have identified certain themes that appear in this video, such as bait tricks that appeal to you and as unorthodox or modify users. So how people might hack these devices
and use for unintended purposes,
people are examine Instagram videos to understand whether there's health warning labels associated with them,
as well as how do these videos explain health effects of E cigarettes?
And nicotine use as well as the marketing content.
These are just some examples of what’s been examined on YouTube videos.
However, there is a lot of limitation in current methods,
so all of these prior studies have used human coding,
which means that you know we have humans going in and and and
watching a video to identify these themes and really limit the number of videos that could be examined. So in these studies they examine about 50 to 350 videos, but in our previous study we examined big trip videos on YouTube. We found that there is like 156,000 videos just on vape tricks along and other studies have found that 2200 new E-cigarette videos are being uploaded every month. So, advances in computational methods can enhance the methods used to
analyze social media data to inform tobacco regulatory science.

So the other issue with social media is that social media custom tailors the content to users.

So we know that there is a lot of cigarette content and this I should say this algorithm of how social media content tailors the users is proprietary and we really don’t know what kind of content user being exposed to,

so understanding the types of content that you would mute or exposed to is really important to inform regulations as well as how to create prevention strategies such as counter marketing.
And no study has yet try to mimic youth conducting the search and then apply machine learning to understand all the data retrieved. So advanced computational methods can be applied to overcome these limitations and gaps, or another limitation is getting more. How do we get these data or videos rapidly so some platforms provide access via application programming interfaces APIs while other platforms require more involved coding to build data scrapers and API’s could potentially deliver thousands or dozens...
even millions of posts per day. And additionally computational methods. Can be used to understand topics related to tobacco prevention using large social media datasets. So now I will sort of switch gear to talk about two studies that we've used to analyze YouTube content on E cigarettes and these studies use unsupervised machine learning, rule based classification, network analysis as well as supervised machine learning. The study one we wanted to understand whether E cigarette content on YouTube differs by U2 youth
NOTE Confidence: 0.97449684
00:44:45.943 --> 00:44:46.857 demographic characteristics.
NOTE Confidence: 0.97449684
00:44:46.860 --> 00:44:48.828 To understand whether you think content
NOTE Confidence: 0.97449684
00:44:48.828 --> 00:44:50.959 is being tailored to certain views.
NOTE Confidence: 0.97449684
00:44:50.960 --> 00:44:52.196 To do this,
NOTE Confidence: 0.97449684
00:44:52.196 --> 00:44:54.668 we create a 16 fictitious viewer
NOTE Confidence: 0.97449684
00:44:54.668 --> 00:44:56.556 profiles and these viewer
NOTE Confidence: 0.97449684
00:44:56.556 --> 00:44:58.886 profiles were separated by age.
NOTE Confidence: 0.97449684
00:44:58.890 --> 00:45:01.590 So 16 year olds and 24 year olds by
NOTE Confidence: 0.97449684
00:45:01.590 --> 00:45:03.817 gender as well as race ethnicity.
NOTE Confidence: 0.97449684
00:45:03.820 --> 00:45:05.000 We may profile for white,
NOTE Confidence: 0.97449684
00:45:05.000 --> 00:45:05.391 black,
NOTE Confidence: 0.97449684
00:45:05.391 --> 00:45:07.737 Hispanic youth and we used factory
NOTE Confidence: 0.97449684
00:45:07.737 --> 00:45:09.750 reset Android phone with Orbot
NOTE Confidence: 0.97449684
00:45:09.750 --> 00:45:11.675 app to delete all personalization
NOTE Confidence: 0.97449684
00:45:11.675 --> 00:45:13.410 based on search results.
NOTE Confidence: 0.97449684
And these are the search results

are words that we use related to E cigarettes and we conducted this search inmate 720.

And we obtain 140 videos which is equivalent to about 7 pages

And so after we remove all the duplicates we had 4201 non duplicate videos in our search result.

The first we wanted to understand, you know we had to develop a cool bug you had to develop a cool bug to understand what we’re examining.

So what we’re interested in examining was like what are the videos being
00:45:48.686 --> 00:45:50.166 related to E cigarettes, right?

00:45:50.166 --> 00:45:51.696 So were they product reviews, vape tricks, health information?

00:45:53.700 --> 00:45:54.700 You know?

00:45:54.700 --> 00:45:57.628 What were these videos talking about?

00:45:59.151 --> 00:46:00.587 And then we want to know who are the people who are uploading these videos,

00:46:00.590 --> 00:46:03.022 where they private users, retailers and we want to know what types of E cigarette products are being featured or the eliquids

00:46:05.130 --> 00:46:10.750 box mod pods and so on.

00:46:08.710 --> 00:46:13.318 We also want to see if there were actually selling these products

00:46:14.930 --> 00:46:17.117 to youth and so we buy we look to
00:46:17.117 --> 00:46:19.235 see whether this external links
NOTE Confidence: 0.97449684
00:46:19.235 --> 00:46:21.495 for purchasing and discount codes.
NOTE Confidence: 0.97449684
00:46:21.500 --> 00:46:23.830 So once we quoted this book, I’ll catbug.
NOTE Confidence: 0.97449684
00:46:23.830 --> 00:46:25.690 We’re two independent reviewers
NOTE Confidence: 0.97449684
00:46:25.690 --> 00:46:28.060 randomly review the finalizer themes,
NOTE Confidence: 0.97449684
00:46:28.060 --> 00:46:31.410 and then we establish integrative
NOTE Confidence: 0.97449684
00:46:31.410 --> 00:46:32.080 reliability.
NOTE Confidence: 0.97449684
00:46:32.080 --> 00:46:34.426 And then after that one quarter
NOTE Confidence: 0.97449684
00:46:34.426 --> 00:46:35.599 labeled 1000 videos,
NOTE Confidence: 0.97449684
00:46:35.600 --> 00:46:38.366 which was used to train supervised
NOTE Confidence: 0.97449684
00:46:38.366 --> 00:46:40.569 machine learning algorithms for study one,
NOTE Confidence: 0.97449684
00:46:40.569 --> 00:46:42.480 I’m going to focus on video themes
NOTE Confidence: 0.97449684
00:46:42.544 --> 00:46:44.488 because our goal was to understand
NOTE Confidence: 0.97449684
00:46:44.488 --> 00:46:46.220 whether the video theme content
NOTE Confidence: 0.97449684
00:46:46.220 --> 00:46:47.740 was different among users.
NOTE Confidence: 0.97449684
00:46:47.740 --> 00:46:48.167 However,
the methods are the same for both studies.

So using network analysis we plotted exposure similarities as a network of demographic attributes and videos.

So what you see here is a graph of male, female and by different age groups and the thickness of this purple line indicates the normal number of common videos.

So what we see that both 24 year old profiles have the most common videos.

And we also use K means clustering, which is a powerful unsupervised machine learning algorithm that finds similarity.
between items and grouped them into clusters without the human input. And then we used human data. A human labeled data as an input to graph convolutional network for machine based classification of the 4201 videos, and we found that just north of high accuracy and using GCN we were able to identify what the video themes were. So 49% of the videos were product reviews, or informational or or modifying these are videos that teaches people how to use an E cigarette or how to modify or hack in E cigarette
00:48:08.470 --> 00:48:10.366 15% or health information.
NOTE Confidence: 0.864424234

00:48:10.366 --> 00:48:13.555 Videos about E cigarettes and 9% were just like other types of videos.
NOTE Confidence: 0.864424234

00:48:18.890 --> 00:48:21.355 And so after performing clustering classification, we calculate the percentage of each video type in each category by demographic groups.
NOTE Confidence: 0.94767824666667

00:48:21.355 --> 00:48:23.011 So what we find here is that.
NOTE Confidence: 0.94767824666667

00:48:23.011 --> 00:48:25.333 The green color is the product of you, so these are videos that talk about you know like give product reviews on the product.
NOTE Confidence: 0.94767824666667

00:48:25.333 --> 00:48:28.044 The product reviews represented by the green color is more common among 24 year old profiles.
NOTE Confidence: 0.94767824666667

00:48:28.044 --> 00:48:31.950 So what we find here is that.
NOTE Confidence: 0.94767824666667

00:48:31.950 --> 00:48:33.798 The green color is the product of you,
NOTE Confidence: 0.94767824666667

00:48:33.800 --> 00:48:35.728 so these are videos that talk about you
NOTE Confidence: 0.94767824666667

00:48:35.728 --> 00:48:39.437 product and we find that the product reviews represented by the green color
NOTE Confidence: 0.94767824666667

00:48:39.437 --> 00:48:41.147 is more common among 24 year old profiles.
NOTE Confidence: 0.94767824666667

00:48:41.147 --> 00:48:44.540 Health health is represented by
NOTE Confidence: 0.94767824666667
Orange is similar or cross a little bit more common among males.
And what you what’s interesting here is that the lighter bluish purplish color here is informational videos where how to use an Instagram or how to modify an Instagram. And that’s a lot more common among underage female group.
And other other videos are more common, represented by the darker purple here for male 16 year olds, which is concerning because these videos had content like you know related to cannabis vaping and other vape tricks and so on.
00:49:21.570 --> 00:49:23.780 So there is concerning content

00:49:23.780 --> 00:49:25.990 that shows that more tailored

00:49:26.068 --> 00:49:28.280 towards younger younger youth.

00:49:28.280 --> 00:49:30.892 So our results show that demographic

00:49:30.892 --> 00:49:33.260 attributes does factor into

00:49:33.260 --> 00:49:35.036 YouTube algorithmic systems.

00:49:35.040 --> 00:49:36.920 In the context of esseker

00:49:36.920 --> 00:49:40.590 related queries on YouTube,

00:49:40.590 --> 00:49:43.270 we found that the similarities between

00:49:43.270 --> 00:49:45.504 exposure for male and female 24 year

00:49:45.504 --> 00:49:47.039 olds and actually higher than than

00:49:47.040 --> 00:49:48.670 the connection between other pairs.

00:49:48.670 --> 00:49:50.872 We also found that underage

00:49:50.872 --> 00:49:52.918 instructional videos on E cigarettes,
while all the age groups were most exposed to product reviews.

So all of this is concerning because.

We because this shows that underage profiles, right so 16 year olds are able to or are exposed to E-cigarette content despite YouTube having policies about prohibiting Easter great content to their underage viewers, such as product reviews.

So now I’ll talk about our second study.

So we identify we have four areas of interest, which is, you know, what are the video themes? Who are the people uploading these videos? You know what types of E-cigarette...
products are being featured and is their presence of sales and discounts.
So what we want to do is we you know we could use human coders to identify them, but we wanted to know can we use supervised machine learning to identify these key areas that could inform E cigarette prevention? So what is machine learning? Machine learning is powerful and it could be used to examine a large data set. So in this case large, many videos machine learning has been used to examine social media content around tobacco use.
However, no studies have examined YouTube videos using machine learning. So this is a quick overview of what a machine learning does, so using an algorithm to predict something. So in this case, if we’re interested in saying you know, can we use machine learning to identify if a video featuring an E cigarette first we need to teach the algorithm what an E cigarette is, right? So we teach it, if it’s jewel, if it’s east, sick, if it’s vape, then it’s considered an E cigarette.
and this is A and this is, this data set is now. Used to train the machine learning algorithm and the algorithm learns from this example data set and later uses a different data set to predict whether they could identify an E cigarette. So if it correctly identify that there is an issue, regret that he’s a successful model. If it fails to identify where an E cigarette exists, when it doesn’t, then we reach train this machine article rhythm until we could.
achieve a successful classification.

So in our study, this is a model performance of our machine learning models for each of the four categories.

F1 score is a measure of test accuracy.

It's calculated from the precision and recall of a test.

And this is a like a pretty good score considering the complexity of the themes that we were identifying.

So what do we find?

So this is a little more detailed look into video themes that we use in this case study versus our study.

One that's what we have more themes here, and we also similarly identify the
00:52:43.544 --> 00:52:45.429 product views were the most common.
00:52:45.430 --> 00:52:47.366 And if you see a picture image here,
00:52:47.370 --> 00:52:48.833 this is an example of what a
00:52:48.833 --> 00:52:50.133 product review look like, right?
00:52:50.133 --> 00:52:52.148 This is Jewel starter Kit
00:52:52.148 --> 00:52:53.357 unboxing and review.
00:52:53.360 --> 00:52:55.918 And we also found that 72nd highest
00:52:55.918 --> 00:53:00.954 video theme was modified video that
00:53:00.954 --> 00:53:03.126 teaches people how to modify and
00:53:03.205 --> 00:53:05.860 informational videos on how to use
00:53:05.860 --> 00:53:08.660 health information was 11% other
00:53:08.660 --> 00:53:08.660 themes that were still ysaguirre.
00:53:08.660 --> 00:53:11.060 9% of marijuana related things
00:53:11.060 --> 00:53:13.616 was 6% and other irrelevant theme
00:53:13.616 --> 00:53:16.010 which is like non E cigarette
00:53:16.010 --> 00:53:16.010
theme for five percent 5.6% and vape chicks was one point 1%.

So product type, so this is all the different types of products that we identified through machine learning and what this actually shows is that there are a variety of different types of E cigarette products that are being featured on YouTube. These are independent users who post almost exclusively about bathing. So when you go to the channel page to see
what kind of videos they've uploaded, it was mostly related to vaping, but they were not directly connected to vaping company, so we cannot verify that their influences or not. So these are some examples of like account of people who've a person. Vape enthusiasts of channel page. As you could see, all the contents related to vaping. This is problematic because when it comes to regulating content, you cannot regulate private users, right? You can’t tell the regular
person to say you know.

Don't post things about vaping.

However, you could regulate influencers who get paid by the industry to post their products and the the. The difficulty with vape enthusiasts is that there's no way to tell who are vape enthusiasts, who are influencers and her regular users. 21% are stores, 12% is other sources and six point 4% of medical community and 6% of private users. So 59% of video did not have any discount or links 34% of the videos had external links for purchasing.
and 5% or have other discount methods

and one point 7% had discount.

So this is a screenshot of instructional videos like beginning

beginners vaping tip that also had

a link that you could purchase

as well as a coupon code.

For purchasing,

So what do we find in this study?

We found that I complicated things

relevant to E cigarettes could be

identified using machine learning and

fictitious youth viewer profiles on YouTube.

We identified videos that violated

YouTube tobacco policy restricting
promotional content to underage minors,
NOTE Confidence: 0.918577972727273
such as product reviews and purchasing links.
NOTE Confidence: 0.918577972727273
Again, there was a high level
NOTE Confidence: 0.918577972727273
of industry presence and such
NOTE Confidence: 0.918577972727273
as faith enthusiast at stores.
NOTE Confidence: 0.816114232
So overall conclusions, you know.
NOTE Confidence: 0.816114232
Mixed methods such as qualitative
NOTE Confidence: 0.816114232
analysis using human labellers and
 computational methods can really reveal
NOTE Confidence: 0.816114232
E cigarette use content to inform youth,
NOTE Confidence: 0.816114232
tobacco prevention and social media has
NOTE Confidence: 0.816114232
really a really rich data and has a good.
NOTE Confidence: 0.816114232
You know you could have a really good
NOTE Confidence: 0.816114232
understanding of youth behaviors as well as
NOTE Confidence: 0.816114232
promotion and sales that youth can access.
NOTE Confidence: 0.816114232
And again, this is our current
And to prevent youth E-cigarette uptake, regulation of social media, a promotion that occurs in social media is really needed. So you know this is one example of how social media could be leveraged using qualitative and computational methods to understand certain behaviors that could prevent KENS. Has cancer prevention implications like tobacco use? But certainly this type of methods could be used to understand other behaviors that
has direct implications to preventing cancer,

such as, you know, physical activity,

diet, obesity as well.

So I’d like to acknowledge our funding stores as well as Yale Tobacco Center of the Study on Tobacco Regulation, tobacco product of Youth in addiction and also our team in University, Texas Austin, who is leading the computational methods.

So thank you for your attention.

Thanks Doctor Kang that was that was great.

And it’s open for questions, please put him in the chat.

I know we only have a few minutes,
but maybe we could stay over for a minute or two.

If people have questions.

Have you? Reached out to YouTube and showed them your data and asked whether they,

I mean it does sound like there’s clear evidence that their policies are being violated.

Presumably they have the computational firepower to be able to do similar things.

Is it something that they may be convinced to look into?

Yeah, that’s a great question.

You know, I have a paper cut currently under review that’s
looking at all of the self imposed.

Social media policy.

Across all the social media platforms

on tobacco and and not surprisingly,

you know all of the social media

They’re not being enforced,

so so hopefully you know this will

Aside from You Tube.

But just looking at all the social media

platforms and what more could be done.

Yeah, and I think could get,

you know,

I think 1 translational component

is that we publish in peer review
00:58:39.169 --> 00:58:40.887 journals and a lot of this information

00:58:40.887 --> 00:58:42.506 don’t get out into the bigger world

00:58:42.506 --> 00:58:44.053 and I think just doing some of

00:58:44.060 --> 00:58:45.728 that legwork might be important in

00:58:45.728 --> 00:58:47.238 getting some of these attention

00:58:47.238 --> 00:58:48.938 for two social media platforms.

00:58:50.980 --> 00:58:53.668 It’s important work.

00:58:53.670 --> 00:58:55.128 So it’s it’s a few minutes

00:58:55.128 --> 00:58:56.771 after the hour doesn’t look like

00:58:56.771 --> 00:58:57.995 there’s any more questions.

00:58:58.000 --> 00:59:00.244 So again, thank you to both

00:59:00.244 --> 00:59:02.421 the the presenters for very

00:59:02.421 --> 00:59:04.218 interesting discussion and.

00:59:04.220 --> 00:59:06.388 We’ll see you at the next grand rounds.

00:59:06.390 --> 00:59:07.260 Thank thank you.
00:59:08.510 --> 00:59:09.170 Thank you.