tactiq.io free youtube transcript

Dr Mia Petljak, Broad Institute of MIT and Harvard, APOBEC-associated mutagenesis

https://www.youtube.com/watch/w25jBc4XNtg

00:00:00.380 [Music]

00:00:10.839 thank you very much so I understand that

00:00:13.849 this is the last session before the

00:00:15.230 lunch I hope I can still grasp your

00:00:16.880 attention for just a little bit longer

00:00:18.220 now the two mutational signatures that I

00:00:22.009 would like to discuss with you today

00:00:23.439 we're first discovered actually in

00:00:25.820 breast cancer in 2012 but it was really

00:00:29.390 in 2013 when the field was astonished by

00:00:32.330 their relevance which was at a time when

00:00:34.550 mutational signatures were for the first

00:00:36.230 time extracted across most cancer types

00:00:39.470 available at a time and these two

00:00:41.870 signature signatures 2 and 13 were found

00:00:44.449 to be highly prevalent and today we

00:00:47.059 think that together the reflects traces

00:00:49.579 of one of the major mutational processes

00:00:51.079 in human cancer as such so why together

00:00:54.770 well obviously these signatures are

00:00:56.960 different because signatures 2 is

00:00:58.940 characterized by c-to-t mutations very

00:01:01.760 signature 13 by C 2 G and C 2 a

00:01:04.250 mutations but they have two things in

00:01:06.590 common so first of all both signatures

00:01:09.140 are characterized by mutations at site

00:01:11.030 designs that being seated T C to G or C

00:01:13.759 to a mutation and second of all if you

00:01:16.280 look at the sequence context cytosines

00:01:18.860 in both signatures 2 and 13 are preceded

00:01:21.950 by a time in base so very specific to

00:01:24.050 nucleotide sequence context now when

00:01:26.899 these were discovered people went into

00:01:28.520 the literature to try to see whether

00:01:30.890 these kind of mutational patterns were

00:01:33.020 reported in any kind of biological

00:01:34.520 context before to try to figure out what

00:01:37.729 is causing them in cancer and the answer

00:01:40.190 was yes they were reported and as a

00:01:42.319 matter of fact it is a probe excited in

00:01:44.750 the amylases that can cause these exact

00:01:47.149 mutation or patterns when they're over

00:01:49.250 expressed in experimental systems in

00:01:51.709 vitro now APOBEC enzymes are a family of

00:01:56.090 10 highly homologous to highly similar

00:01:58.250 enzymes that in humans actually evolved

00:02:00.920 only recently so duplication of the

00:02:03.140 ancestral gene on the chromosome 22 so

00:02:05.509 their position on chromosome 22 1x2

00:02:07.789 other and they're very similar and all

00:02:10.008 of them shared is highly homologous all

00:02:11.569 societies in the ammonites domain now

00:02:15.560 the interesting thing here was that we

00:02:17.959 didn't actually know about these things

00:02:19.790 from cancer we knew about them from

00:02:21.950 immunity because as it

00:02:24.230 loud arabic demonizes although they can

00:02:27.400 mutate its DNA in vitro they were mainly

00:02:30.739 known to us from mutating RNA and from

00:02:33.409 mutating RNA of retroviruses and

00:02:36.010 retrotransposons because in fact they

00:02:38.299 are part of the innate immune response

00:02:39.700 so the context where these were studied

00:02:42.349 the most before was HIV and in other

00:02:45.830 words upon infection these are meant to

00:02:48.560 act on the viral sequence to edit it and

00:02:51.019 to stop the virus from replicating so

00:02:53.930 this really left the field in wondering

00:02:57.079 is it adult possible that something that

00:02:59.510 actually evolves to protect us from

00:03:01.579 viral infection then also contributes to

00:03:04.340 the development of cancer and as such is

00:03:06.769 a matter of fact a double-edged sword

00:03:09.290 so this triggered the whole new field of

00:03:14.090 studies trying to link up effects and

00:03:17.060 mutations observed in cancer and it is

00:03:19.430 only sufficient actually if you go into

00:03:20.870 PubMed and you put in Apple back and you

00:03:22.670 put in cancer so two terms together you

00:03:25.129 will see the flood of studies of these

00:03:27.319 two fields colliding that started around

00:03:29.540 the time that I just described to you

00:03:30.980 and continues up until today so fast

00:03:34.130 forward to today to 2019 when the

00:03:36.440 algorithms that we have to detect

00:03:37.910 mutation or signatures are a lot more

00:03:39.739 refined and what I'm showing you here is

00:03:42.319 just a list of samples where signature

00:03:44.239 to and signature 13 are found and very

00:03:46.910 easily you can say well it's a lot and

00:03:48.620 if you were to quantify it it's actually

00:03:50.480 almost 80% of all cancer types in which

00:03:53.599 we find these signatures and then it is

00:03:56.090 not that these signatures contribute to

00:03:58.069 small proportion of individual cancer

00:04:00.139 samples from a particular cancer type as

00:04:02.450 a matter of fact they contribute to more

00:04:04.970 than a third of individual cancers from

00:04:06.950 some of the major types affected so we

00:04:09.230 heard about it Dhafir goose lung breast

00:04:11.480 head-and-neck bladder surrogate cervix

00:04:13.910 and many many other and then you can

00:04:16.370 also look at the individual numbers of

00:04:18.440 mutations that these signatures

00:04:19.608 contribute to individual cancers and

00:04:21.380 here each cancer is a dot it belongs to

00:04:24.229 a specific cancer type and on the y-axis

00:04:26.930 you have numbers of mutations actually B

00:04:29.479 to Twitter signature 2 or 13 so this is

00:04:31.550 taken from actually cosmic and you can

00:04:33.830 see where as in some cases these

00:04:35.210 signatures can contribute small burden

00:04:37.129 mutations most commonly actually they

00:04:40.429 contributes hi-hi mutational load so

00:04:43.039 underline mutational process is the

00:04:44.839 hyper mutational one now if you were now

00:04:49.339 to actually compare these two lists even

00:04:51.860 though they're ordered in a different

00:04:53.029 way they're actually the same cancer

00:04:54.349 types and as a matter of fact there are

00:04:56.479 same cancer samples because you will

00:04:59.149 find signature two and signature 13 most

00:05:02.149 commonly in all cancer samples here

00:05:04.459 albeit a little bit at a different

00:05:06.619 proportion and why is that comes down to

00:05:09.379 thinking a little bit about the

00:05:10.519 mechanistic of how these arise and what

00:05:12.919 we think is happening so the process

00:05:14.959 starts by aqua Beck's one of them

00:05:18.069 deaminating site designs which then goes

00:05:20.300 into uracil and then depending on how

00:05:22.819 you repair or how you act on that uracil

00:05:24.889 you will either get a C 2 T mutation so

00:05:27.259 this is a upon replication or a C - G

00:05:30.709 mutation and I don't know what happened

00:05:32.360 to my slides regardless there are two

00:05:34.729 different mechanisms and depending on

00:05:36.529 the direction which you go into you can

00:05:37.909 get a C - Tim mutation or a c2 n a C - G

00:05:40.969 mutation

00:05:43.129 now most of the signatures that we

00:05:44.990 talked about today including signatures

00:05:46.669 2 and 13 are genome-wide and so here I'm

00:05:50.629 showing you proportion of the chromosome

00:05:52.849 X and you have mutations positioned

00:05:56.029 alongside that portion of chromosome X

00:05:58.490 they're colored based on the base

00:06:00.469 substitution and then the Y X is what is

00:06:03.079 plotted is inter mutational distance so

00:06:05.539 mutational distance between consecutive

00:06:07.579 mutations and what you can see is that

00:06:10.279 they are more or less equal distant this

00:06:13.699 is what we see for the majority of the

00:06:15.499 signatures however every now and then

00:06:18.379 this is observed and first observed in

00:06:21.229 the breast cancer actually we get these

00:06:23.240 clusters of mutations that are referred

00:06:25.339 to as mutational showers or cottages by

00:06:27.349 the Greek word thunderstorm and then you

00:06:30.169 can take this individual dots so there

00:06:32.179 are mainly c 2 t mutations but there are

00:06:33.769 others that you can't see because of the

00:06:35.269 high burden of C - T's and then you can

00:06:37.069 plot these into the sequence context and

00:06:39.319 what you get out is actually again

00:06:41.689 signatures 2 and 13 like patterns so

00:06:44.869 again we think it is a Quebec that is

00:06:46.849 behind this kind of clustered mutational

00:06:49.279 signatures and indeed if you

00:06:50.480 do the experiment in East if you

00:06:51.800 overexpress it he will see that but

00:06:53.720 obviously the process is different

00:06:55.070 because it's not a genome-wide random

00:06:57.290 acquisition of mutations throughout all

00:06:58.970 of the human chromosomes here our probe

00:07:00.890 acts are potentially localized to one

00:07:03.290 specific spot now for the second part of

00:07:07.550 my talk I'd like to actually talk about

00:07:09.790 links that were made in the field

00:07:12.620 between up of X and mutations in cancer

00:07:14.510 within three main streams of research

00:07:16.700 and that is experiment experimental sort

00:07:19.760 of evidence from genetics from

00:07:21.080 mutational features and expression in

00:07:22.970 cancer so I will start by mutational

00:07:26.000 features because this is how the link

00:07:27.560 was originally a made based on the

00:07:29.540 literature and then you can actually go

00:07:31.640 and confirm this in the lab so you can

00:07:33.140 take each one of these Apple Beck

00:07:34.490 members you can overexpress them in

00:07:36.410 yeast and you will hope to find these

00:07:38.330 particular patterns or mutations that we

00:07:40.220 see in cancer to tell abilities this

00:07:41.810 Apple baby that is likely causing

00:07:43.640 mutations now based on these experiments

00:07:46.580 very early on so the members can be

00:07:48.290 dismissed because actually when you

00:07:49.520 overexpress and they don't cause

00:07:50.600 mutations so these two are unlikely to

00:07:53.060 be our mutaters for others you can take

00:07:56.900 so this is just an experiment showing

00:07:58.310 you that over expression for Apple Beck

00:08:00.140 3G so you can take all of the mutations

00:08:02.570 adjusted cytosine bases because our

00:08:04.100 signatures are characterized by

00:08:05.270 mutations in cytosine bases and then you

00:08:07.700 can look at the sequence context five

00:08:10.100 days is left five-prime and three-prime

00:08:11.800 to the mutated base and what we are

00:08:14.870 hoping to see is that the base at minus

00:08:17.180 one position is red so it's a T because

00:08:20.660 it is this is the pattern of our

00:08:21.950 signature so for Apple Beck 3G obviously

00:08:24.500 this is not the case there's a lot of

00:08:25.970 blue at the position minus 1 these are

00:08:27.680 C's so these two is likely not our

00:08:30.170 candidate others however and here I'm

00:08:32.599 showing you only some of them turn out

00:08:34.490 to be inducing mutations in exactly the

00:08:36.650 sequence context that we need and you

00:08:38.240 can see the predominance of T mutations

00:08:40.130 at the minus 1 position now in so from

00:08:44.270 say from this sort of experiment you

00:08:46.490 could say all of these are potential

00:08:48.410 mutaters now in 2013 3 independent

00:08:52.240 expression based studies came out where

00:08:55.040 people basically took those cancers with

00:08:56.780 a lot of signatures 2013 and they try to

00:08:59.150 associate burdens of these signatures

00:09:01.400 with

00:09:02.259 expression of individual APOBEC members

00:09:04.600 and what came out of these three studies

00:09:06.759 mainly is the littles expression of the

00:09:08.919 apobec3 that was best associated with

00:09:11.709 the burden so here it is a potential

00:09:13.929 mutator however one thing to remember

00:09:16.389 from this kind of studies is that when

00:09:18.309 you look at RNA and DNA product from

00:09:20.259 whatever kind of sample or a cell in DNA

00:09:23.679 we measure our signatures in RNA we

00:09:25.600 measure expression of the genes our

00:09:27.519 signatures required potentially

00:09:29.919 throughout lifetime from fertilized egg

00:09:32.289 whereas expression is self state

00:09:34.749 captured at that time point so

00:09:36.249 potentially what you're doing is

00:09:37.749 associating two different time points

00:09:40.589 and then as it happens actually probably

00:09:43.689 and what I would say one of the most

00:09:46.689 stringent evidence for the involvement

00:09:48.549 of our products in human cancer comes

00:09:50.350 from genetics because there is a common

00:09:52.809 germline polymorphism that is

00:09:55.269 essentially deletion of majority of the

00:09:57.489 upper back 3b sequence I've told you

00:09:59.229 that these things exists together one

00:10:01.809 next to other on chromosome 22

00:10:03.639 and this germline polymorphism generally

00:10:06.189 meaning is inherited so it's every in

00:10:07.869 every cell of those people who inherited

00:10:09.639 will delete Apple back 3b and those

00:10:12.819 people who have the germline

00:10:14.019 polymorphism have increased risk of

00:10:16.689 breast cancer and those breast cancers

00:10:19.179 usually have increased burdens of

00:10:20.829 signatures 2 and 13 and this tells you

00:10:23.169 two really important things one of which

00:10:25.539 is now you have a genetic evidence for

00:10:27.339 development of effects in human cancer

00:10:29.619 so this is not an artificial model this

00:10:32.109 is human data and second of all is that

00:10:34.749 you may not need or you don't need in

00:10:36.639 this case f of X 3 B and as a matter of

00:10:39.609 fact there is a mechanism through which

00:10:41.649 you can explain this because there is a

00:10:43.829 promoter or enhancer region of f of X 3

00:10:46.959 bitter if it comes down to up above 3a

00:10:48.639 and we think is drives over expression

00:10:50.859 of upper back 3a so this really cause

00:10:53.559 kind of chair you know there was a shift

00:10:55.629 in the gears in the field from f of X 3

00:10:57.579 B to upper back 3a and as it happens

00:11:02.199 over the past year is actually our

00:11:04.029 pattern based studies also evolved and

00:11:06.879 what people have realized that even

00:11:09.039 though apobec3 a and apobec3 B when you

00:11:11.739 over express them in yeast

00:11:13.580 the patterns used look quite similar if

00:11:16.279 you look more closely to essentially

00:11:18.320 repeat the experiment and you look a

00:11:19.910 little bit closer at the position - - so

00:11:22.580 next to that timing they actually have a

00:11:25.100 slightly different sequence preference

00:11:27.080 so Aquabat 3a prefers purine bases where

00:11:31.399 is Apple back three B so f of X 3a

00:11:33.649 prefers pyrimidine bases whereas Apple

00:11:35.510 vector EB prefers purine bases and then

00:11:38.540 because now you can differentiate

00:11:39.800 between two signatures the Disco's in

00:11:42.200 yeast you can also go into human cancers

00:11:44.209 and as well what do they reflect and as

00:11:46.760 shown here for bladder breast head and

00:11:48.620 neck and lung you can see that there is

00:11:50.329 a preference for the pure for the

00:11:52.160 pyrimidine base and thus again further

00:11:54.200 evidence that is apobec3 a that is the

00:11:56.420 main mutator in these cancers now here

00:12:00.320 is some data that I wasn't originally

00:12:03.290 going to show but there was a paper that

00:12:04.790 actually came out in science about a

00:12:06.560 couple of weeks ago where people have

00:12:08.810 taken a further extended sequence

00:12:11.240 context so you can go further so our

00:12:13.640 signatures are actually nucleotide bases

00:12:15.200 I've just shown you what happens if you

00:12:16.579 take one more in this particular example

00:12:18.079 you can go further and they have

00:12:20.990 analyzed this kind of a public specific

00:12:23.240 mutation so cytosines when preceded by

00:12:25.430 a--they and what they have found is that

00:12:27.920 actually in some of the upper back what

00:12:30.290 we think are a public hotspots they

00:12:32.480 appear in the sequence context that's

00:12:34.459 predicted to form these loops and then

00:12:37.160 you can again take your app of x3a you

00:12:38.930 can again take your upper-back 3 b you

00:12:40.550 can express them in eastern you can as

00:12:41.990 well which one is causing mutations and

00:12:43.610 loops and as it turns out again it is up

00:12:46.070 about 3a so what I have just told you is

00:12:49.880 the sort of ongoing debate that's going

00:12:51.770 in the field but there is now growing

00:12:53.120 and growing evidence about the

00:12:54.529 involvement of alphabet 3a from genetics

00:12:56.720 from mutational features expression

00:12:59.480 based studies point to Apple back three

00:13:01.070 but I'll discuss the potential

00:13:02.180 limitation of those that is not to

00:13:03.649 dismiss apobec3 a but it is to say that

00:13:06.380 from this other studies apobec3 maybe

00:13:08.449 that major hyper mutator

00:13:10.130 some of them we dismissed and some of

00:13:12.890 them actually can't cause mutations that

00:13:14.570 are sequence context of interest but

00:13:17.329 actually they don't have access to

00:13:18.890 nucleus so they don't have access to DNA

00:13:20.420 or they have a tissue specific

00:13:22.570 expression that doesn't reflect cancer

00:13:24.770 is very see these signatures

00:13:26.310 now one stream of research that is

00:13:29.440 missing from the field is a direct

00:13:31.800 experimental evidence so everything that

00:13:34.180 I've talked about so far are links and

00:13:36.100 our associations what we still don't

00:13:38.770 know whether it is really a babies that

00:13:41.080 are in a human cancer cell generating

00:13:42.970 this signature we have links we don't

00:13:44.530 have the direct evidence which Apple

00:13:46.660 Beck is it good evidence for 3a but is

00:13:48.970 it and then finally and most importantly

00:13:51.640 what is actually activating these things

00:13:53.200 so what is making these things go every

00:13:54.790 now and then crazy and mutate human DNA

00:13:57.580 instead of viral RNA now the reason that

00:14:03.640 there is no director spend on

00:14:05.290 experimental evidence is because the

00:14:07.660 field has been missing models and the

00:14:09.400 appropriate model in this case would be

00:14:11.260 a human cancer cell that endogenously so

00:14:13.750 without any kind of experimental petrol

00:14:15.430 Bayesian acquires this signature over

00:14:17.590 time so that is the condition number one

00:14:19.330 and then the condition number two is

00:14:21.070 that it is genetically amenable so you

00:14:22.720 can delete top of X and you can try to

00:14:24.760 stop mutation acquisition so in 2019 or

00:14:28.600 actually earlier this year we have

00:14:31.000 annotated mutational signatures across

00:14:33.370 thousand and one human cancer cell lines

00:14:35.290 so this is the largest panel of cell

00:14:38.260 lines more and includes most cell lines

00:14:40.690 used viola in cancer research alongside

00:14:43.150 smaller proportion of pdx models that we

00:14:45.070 looked at as well and we were actually

00:14:46.930 able to tell which of these thousand and

00:14:48.760 one cell lines has which panel of

00:14:50.200 signatures now once we know that the

00:14:54.130 second requirement for a good model is

00:14:55.960 that these signatures continue to be

00:14:57.520 acquired over time so we have actually

00:14:59.470 designed in vitro experiments where we

00:15:01.780 were able to in cancer cell lines with

00:15:03.820 known mutational signatures that we

00:15:05.560 selected from this large panel track

00:15:07.960 mutation acquisition over very specific

00:15:10.300 time frames and then once we were able

00:15:12.640 to do that we were able to account

00:15:14.530 mutations acquired we were able to tell

00:15:16.570 which signature do they belong to and we

00:15:19.090 were actually able to track activities

00:15:21.010 of underlying mutational processes so we

00:15:22.900 were able to tell if the mutational

00:15:25.240 process continues is it discontinued or

00:15:27.550 is it perhaps episodic so this continued

00:15:30.550 you can think about for example if a

00:15:32.500 cell and has a lot of UV light signature

00:15:34.270 it was perhaps derived from a melanoma

00:15:36.340 of a cancer patient who was exposed a

00:15:38.170 lot of you relied but

00:15:39.250 signature will not be ongoing in culture

00:15:40.959 because you're not exposing cells to UV

00:15:42.970 light so this was a large screen over as

00:15:47.379 many signatures as we could find and

00:15:49.240 dozens of sellings and a hundred of

00:15:50.649 clones and I'm not going to go into all

00:15:52.540 of them rather I will just focus on two

00:15:55.449 of our signatures that are upper-back

00:15:56.860 associated and here just for the

00:15:59.050 simplicity I will show you two sellings

00:16:01.209 that obviously have these patterns and

00:16:02.860 how we track them and tracked mutation

00:16:05.230 acquisition over three different time

00:16:06.699 points and the first thing that we saw

00:16:08.829 was actually the DS signatures were not

00:16:10.870 ongoing they disappeared over time in

00:16:12.490 culture and that made perfect sense

00:16:14.079 because in our cultures there is no

00:16:15.939 immune system our public's are known to

00:16:17.470 be part of the immune system so there is

00:16:19.509 nothing to activate them however as we

00:16:22.389 included more and more samples we

00:16:23.980 actually started finding cell lines that

00:16:25.629 do in fact acquire a public signatures

00:16:28.149 over time and this then told us two

00:16:30.730 things first was that actually APOBEC

00:16:33.040 mutagenesis can be initiated in the

00:16:35.290 absence of the immune system so is

00:16:36.850 likely endogenous here and second of all
00:16:39.490 the thing that was noticed was that the
00:16:41.589 mutation rates were highly different so
00:16:43.360 these two for example lineages from the
00:16:45.850 same cell and were propagated for the
00:16:47.379 same number of days around 70 and they
00:16:49.480 acquired very different numbers of
00:16:50.860 mutations and you can see that
00:16:52.180 everywhere for these signatures but it's
00:16:54.550 something that we didn't see for other
00:16:56.050 signatures examined so when we went to
00:16:58.600 look at that more closely we have taken
00:17:00.519 a cell line with signatures 2 and 13 and
00:17:03.040 we have grown it over very short periods
00:17:05.319 of time and we have seen that these
00:17:07.209 signatures disappear and then they come
00:17:09.760 back on because before they disappear
00:17:12.099 again and this really tells you that the
00:17:14.980 mutational processes behind generating
00:17:17.230 these signatures is episodic so
00:17:19.270 presumably whatever is activating aqua
00:17:21.189 Beck's comes in the intermittent bursts
00:17:23.679 of the activity and again I emphasize is
00:17:25.720 something that we haven't seen for any
00:17:27.039 other signatures now importantly what
00:17:31.179 comes out of this study is that now you

00:17:32.860 do have a model where you can actually

00:17:34.450 go on knockouts or remove delete

00:17:37.179 individual alphabets and try to provide

00:17:39.640 a directory spend experimental link for

00:17:41.770 their role in cancer and this is exactly

00:17:44.380 what we are pursuing so here I'm showing

00:17:45.909 you a bt 474 it's a human breast cancer

00:17:48.880 cell and this is a wild-type cell line

00:17:51.370 so it obviously has the signature

00:17:52.870 after 60 days of in-vitro propagation it

00:17:55.150 continues to acquire it and you can see

00:17:57.040 different numbers acquired presumably

00:17:58.930 because of the episodic nature then we

00:18:01.870 can remove up about three B and as it

00:18:04.930 happens the signature keeps on going on

00:18:06.990 however if we remove upper-back 3a we

00:18:10.570 can deplete the signature acquisition

00:18:12.550 and what this now tells you is first of

00:18:15.280 all for the first time you do have an

00:18:17.140 experimental link direct run for up of

00:18:19.750 exonerating signatures in cancer the

00:18:21.880 models works or the system works and

00:18:24.100 third of all in this particular breast

00:18:26.320 cancer cell line up ho-bag 3b is

00:18:28.540 dispensable and this is actually a great

00:18:30.850 collaboration with John my job skied msk

00:18:33.250 CC and we have generated many of the

00:18:35.140 APOBEC knockouts

00:18:36.130 many of the downstream enzymes involved

00:18:38.380 in the mechanistic behind generating

00:18:40.300 different types of the mutations and

00:18:42.040 site designs and and we continue to do

00:18:45.580 that so I would like to thank to Mike to

00:18:48.250 Ludmila to many many great colleagues at

00:18:50.110 Sanger and elsewhere thank you very much

00:18:56.659 you