Label-free AI-enabled Imaging Flow Cytometry for Discovery of Subpopulations of Extravillous Trophoblast from a Stem Cell-based Model of Trophoblast Differentiation

Louise Laurent: I'm Louise Laurent. I'm a professor in the Department of OB/GYN Reproductive Sciences. Also one of the co-chairs for CORI, so thank you very much all for coming to the symposium. It's great to see so many people later in the afternoon. This is Yu-Hwa Lo, my partner in crime for this project. Just wanted to give a pitch for persistence. I was looking back in my emails, and Yu-Hwa and I met in 2017, and we got a GEM grant together, along with Drew Hall, who's also here today.

He presented this morning with Irene. After that, several years later, we got an R21, and then we put in several other grants over the years that were not funded. Finally, now actually Drew and I are in negotiation for an ARPA-H, so persistent, gestational age of nine years for our collaboration is starting to pay off. Hopefully, CORI infrastructure can help people make that process a little bit faster than ours. I think we were pretty junior when we started the collaboration, and now we have gray hair. This is going to be completely different.

This is a collaboration having to do with one of UH technologies and actually trophoblast stem cells in MyLab. I'll jump in the middle to talk about the actual cell system that we are studying together. Yu-Hwa is going to start with talking a lot about the exciting nature of the technology itself. I think he also almost broke the file upload system. I think we have the most megabytes size of, yes, that's true, in the PowerPoint of all the talks today. Please, enjoy the beautiful figures.

Yu-Hwa Lo: Hi. Good afternoon, everyone. Also thank Louise for the invitation for me to have the opportunity to talk to this audience. You can clearly see from the title that this is a close collaboration between medicine and engineering. We set the record of the longest title and probably not very comprehensible. I won't bother you with this title and explain what it means. Also, I try my very best to avoid all the engineering jargons, but if I still couldn't get rid of all of them, I hope that you will forgive me in going through all this jargon.

But what I actually tried to do is we actually try to develop a new and enabling biomedical devices and systems to accelerate biomedical research, discovery, and therapeutic development. That's what we try to do. We look at actually the basic unit for all organisms, that's cells, and we try to actually develop a cell analysis tools to understand and isolate the cells for a deeper analysis of the cells. The far fetched dream is that, can we actually look at the cells and understand its genome, understand its protenamics, understand its metabolics.

Those are very far fetched dreams, but we hope that we could develop some tools that give you a chance to look into that direction aided by the very fast development of AI technology. That's what we try to do. The first few slides, we're going to actually go through a very quick introduction about why we come up with these tools. What you see here on the left is a microscope. On the right is a flow cytometer of FACS fluoresces-activated cell sorting. Microscope gives us a beautiful image of every single cells.

It has all the information, morphological features, phenotypical features of the cells, but lacks one functions that is very desirable and important. That is, can we isolate the cell on particular interest and take it out from the sample and do the downstream molecular analysis? It's very difficult to do it. We have to use microlaser dissection that is a very slow process and damage the cells most of the time. You can handle maybe a few cells in research lab, but certainly not suitable for clinicals. Application. On the right, you see a fluoresces-activated cell sorter.

The good thing is you suspend the cells in the fluid, and you flow the cells in one single fire, in one single land, you use high speed laser to interrogate every single of them. Then when you find some self interest, you isolate them, you trigger the actuator, and you isolate them. You could actually handle cells thousands or even tens of thousands cells per second, so that's very good. But the problem is the information you can handle is very limited, maybe by some biomarkers. For more and more applications, people don't want to disrupt themselves and label the cells with biomarkers.

In more applications, we don't even know the proper biomarker. We still try to discover them. That's the limitation. Our thought from the engineering perspective is, can we combine the two technologies? We have the information content from microscopy, but we have the ability to isolate the self-interest from the fact system. What we try to do is we actually have a high throughput spatial information and then aid it with AI, all three together. Quickly, I will actually just show you how we create this imaging. Remember, the cell is traveling at 20 centimeters per second to one meter per second, and we have to take the image.

Here, you will see that we have a very fast resting scanning lasers. As a cell travels through, we scan through, we project actual image, just like we use the light to cast a shadow of the object. Then once we have the image, we place that on the distribution plot, and we could actually identify the area of interest. Then we will isolate the cell from the area of interest, which we call gating process. Now, how do we isolate? We put the cells in the microflic chips that is 200 micron wide and 70 micron deep.

The cell flows through the microfluidic chamber. When we see the cells of interest, we activate the actuator, which is integrated on chip. This PS electric actuator will be able to push or pull the cells into the particular designated channels. This is how we detect, we image, and we isolate the cells of interest. I will not actually go through the details of how the optical system is designed. Here, I'd just like to actually show you three things. The three images. The system altogether can create nine colored images based on biomarkers, but importantly, three images without labeling.

One is actually on the lower right is what we call backscattered image. It's a reflected light. The reflected light, you see an image. Now, in front, you will see a bright view image, that's transmission images. Then on the far left is actually the forwardscattered image, also called dark field images. This is actually image they collect the small angle scattering. If the cell contains some particles from micro-organelle that can scatter light with a small angle. It doesn't bounce the light back, but it's actually scattered light with a small angle, between 5-12 degrees.

We were going to collect this light and create an image. You have three images representing the property of the cell. We're going to use these three image primarily to determine the tropoblast, the cells from placenta. Here, just quickly, actually a few examples. Here, we try to use this system to the cell-cell interaction. T cell cancer cell actually B cells interaction. The blue one is actually B cells, the red one color is actually the T cells. The green one is actually lymphocyte function associated antigen, so that basically says how the T cell and B cell junction actually bind together.

What's important is actually the two column on the right that shows how the two cells without any labeling, without actually looking at all the fluorescent colors, the two cells are joined together. Here's another example to see T cells binding to the tumor cells while try specific antibody. You can see that all these images, even though the particle are traveling or the cells are traveling at 20 centimeters per second to one meter per second. But here, from now on, we're going to focus on the label-free image, which is a back scattering imaging reflection, forward scattering imaging, and transmission imaging, these three images.

The first example is a stem cell separation, IPSC cell separation. We want to isolate the non-differentiated cells from the differentiated cells. Here you will see that from this image, with a good cell biologist, believe it or not, they could actually put the gate, eventually get very high purity. After sorting, the dark blue bar shows the percentage of undifferentiated cells is almost

100% from the original population. Also, importantly, the process didn't affect the cell viability. Before and after sorting, viability close to 90 or 100%. But then the problem is, the cell are more and more alike, biologists, pathologists cannot tell the difference anymore.

Now, we change the brand of the system from human to AI. That's the central processing. It's an AI processor. Later on, we're going to explain how the AI process is working doing the AI inferencing. To differentiate the cells, to discover the new cell type, and to make the decision whether this is a cell of interest is all done by AI. This is actually the standard network architecture of the AI, we call convolutional neural net. It's a classical neural network because we use imaging. We call the UNet.

If you look at the architecture, it looks like letter U. That's why we call UNet and it certainly have two branches. On the left, we call encoder. On the right we call decoder. On the left part, we receive the input images, and then we turn that into features, just like we look at the person, and at the output of that encoder is a person's height, weight, skin color, pupil color, age, gender, and so on. This becomes features. How many features? We collect a thousand, 24 features, and use it to classify the cells.

Now, the right branch is called the decoder. The decoder is doing just the opposite. It actually takes the features and generate an image. Now, if you look at the image on the right, it's generated by AI. The image on the left is a ground truth, it's input image. How the network is trained is to make sure the image generated by AI is as close to the original image as possible. That's one aspect. The other aspect is actually the classification need to be classified correctly. Now we actually go through some typical exercise.

We put some human white blood cells and tell the AI, we like to actually classify into three groups. There are three groups, they are lymphocyte, the monocyte, and granulocytes. Indeed, they did the 99.9% accurate. But this is something any pathologist can easily do. The next thing is we do unsupervised learning. We train the AI with actually the normal cells purchased from blood bank, and now we actually bring the patient's cell with leukemia. AI said, there's something they hadn't been trained for recognized as blue in this chart.

Then we isolate these cells and actually pathologists look into the cells, it turns out they are the leukemia cells. This is called unsupervised learning and also has very high accuracy of 90 something percent. But this is still something the pathologist can do because pathologist can tell leukemia cells from regular healthy white blood cells. We do something the human cannot do, that is protein translocation. If we actually put the drug to the cells, and that cause protein translocation, in this case, is glucocorticoid receptor. Of course, no image can see protein, but then in the complicated manner, the protein expression level and also its translocation will cause some corresponding morphological change in the features, which human cannot recognize.

But with actually the feature space, over a thousand dimensions, the AI can recognize. Now, AI said that I can divide into two groups. One respond to the drug, the other didn't respond to the drug due to the protein translocation, and they get 85%. This is actually superhuman because without labeling of the protein of target, people cannot do. With that, we'd like to actually have a deeper dive into what's actually in the heart of the system. We actually study four or five different architecture of AI, different models.

Many of you actually probably heard that. But here we like to actually just spend time on one desk we call GAN network. It's generative adversarial network. The network actually have two components. It's very intuitive. One is called discriminator, one is called a generator. The discriminator receive two types of inputs. One is a real image input from the instrument, the hardware. The other is a fake input generated by the AI. The discriminator is trained to the extent that they can tell the real one from the fake one.

Once discriminator is trained, now back to the generator. The generator is actually trying to generate the image that can finally confuse the discriminator to the point that discriminator can no longer tell which one is true or which one is fake. By doing that, the generator received the

training. Once the generator is trained, now the generator is going to output the features. In this case, we actually specify condense to 128 features. Here is actually more detail. I wouldn't actually talk too much about that. One, we use UNet, the other we use CNN.

One thing actually may be helpful for the later part of the discussion is for the generator. We received bright field image, and we generate reflection or call back scattering images. Now, the discriminator will receive the input from the generator. That is one real image bright field, one fake image generated by the generator. But they also receive two real images, the both real bright view image and the real back scattering images. The discriminator tried to distinguish them, and finally, the generator generate such a good quality images to the point that discriminator can no longer tell the difference.

So that's what we try to do, and here, finally, once the generator achieve that capability and we actually create the center part, that's the features, these are the true cell features extracted from the AI network. I will skip this one, skip this, and then, Louise, you can talk about biology now.

Louise Laurent: With all that magic being done, so from my group, we generated cells to put into the system, and from my questions this morning, you guys might have figured out that I have this B in my bonnet about heterogeneity among different biological systems, and in this case, heterogeneity in cells. So I'll just give a little background on the cell type that we're talking about. So these are very recently discovered or derived cell type. These are human trophoblast stem cells. They are derived from the early embryo or the early placenta, and basically, they represent the progenitor or parent cells of all of the sort of placenta unique epithelial cells, and the two that we really care about well.

The human trophoblast cell basically represents this category of trophoblasts, cytotrophoblast CTB. Again, these are the progenitor cells that can self renew, and they can also differentiate into two different sublineages. One is called the syncytiotrophoblast or STB, and these are the cells that transport nutrients and waste products between the mom and the fetus in the placenta, and they also produce pregnancy hormones, like HCG, the thing that's in the urine dipsy that comes from the STB. The other cell type is called the extravillous trophoblast, the EVT, which are less well known.

There's much fewer of them. They are actually even more exciting than the STB. They do so many different things. They anchor the placenta to the uterine wall, the inner uterine wall. They actually remodel uterine blood vessels by invading into them, getting rid of the muscle cells around the vessels. They replace the endothelium, and this is to draw more blood to the placenta during pregnancy, and then finally, they modulate the maternal immune system so that the mom's immune system does not reject the placenta and the fetus.

These are really exciting cells, and so these are the extravillous trophoblast in this area here. What we'll see is that what we're doing with Yu-Hwa is to see what the heterogeneity is both within the CTB state and the EVT state, and this is just a slide to show and hopefully convince you that we know what we're doing. So these are the undifferentiated HTSCs. They express EGFR, or EGF receptor highly, but they don't express the HLAG which is the EVT specific service marker. HLAG is actually non classical MHC molecule and is partially responsible for that immunomodulatory effect, and then after we do this differentiation process, we get EVT.

You can see morphologically, they look very different on the dish, and then they express both EGFR and HLAG so they're in the upper right quadrant of the scattered plot. These are the cell types that we provided to Yu-Hwa lab, and there's like this mad rush the morning of an experiment where we have the cells and we run them over so they don't die on the way and lift them and then put them on disorder. So back to you, Yu-Hwa.

Yu-Hwa Lo: Thank you, Louise. We find that one big problem for challenge for this collaboration is UCSD campus is so huge. [Laughter] We have to carry the cell and run actually to my lab so that actually turns out to be a challenge to keep the cell alive. This is actually.

Louise Laurent: I love [inaudible] it's very difficult to get [Laughter]

Yu-Hwa Lo: It turns out to be very hard, truly. Here, this is actually the workflow. We receive two 6-well-plate. The first one we use to train the AI model. The second one we want the AI system to identify the new cell type within the CTB group, and then to do the downstream analysis, we're going to isolate each one of that. The first group we actually put the 1.5 mils of vial and go through the system without sorting and imaging every one of them. After we image about 100,000 cells, the model is established.

Remember, the model include every cell we register 128 features, but then human can only see XY, so we projected onto an XY plan, and then we color coded them, so you can see the model actually classify the cells into different groups coded by the color. You can see that many of the groups appear to be overlapping because they are overlapping in the two dimensional space. But once you actually have 128 dimensions, they are all separate. Now, the second group, we actually isolate the cells in each of the subtypes.

Each of the subgroups, in this case, into six tubes, and then we send the sample, do the RNA extraction, finishing the library, and then we do the RNA sequencing. As we speak, we're still waiting for the RNA data, so we are not able to.

Louise Laurent: This morning of course. The long again told us.

Yu-Hwa Lo: I'm not going to see that, but this is actually the example. You will see many dots and also color coded, and we divide that the model divide that into nine groups labeled from zero to number 8. Every single dot, if you click, it will show the image of that very cell. Then we will see Group 0 here it turns out to be a doublet, and Group 3 here is a cluster, and Group 2 is another type of mixture of the cells. They are not truly an image of single cells.

The model actually separate them into three separate groups. We know these three groups, we are not interested in that. Then the remaining one is 1, 4, 5, 6, 7, 8, will be the one we're going to sort out. We exclude these three and then focus on the other six, and this actually shows the six groups. For each group, we isolate between 3,500 cells to 10,000 cells, and then we extract RNA through the cubit, and also by analyzer. We actually get a good quality. All their score is above nine, between nine and 10, so they are good quality, and importantly, the cell viability throughout the whole process is above 90%.

In the process, it takes altogether about four hours experiment. Every 30 minutes, we do a live test assays throughout the process using a microfluid chip we develop. So in two minutes, we could actually tell the percentage of the live cells, okay, to monitor any cell degradation in the process, and this basically shows the six groups, and a few examples of the image we get, we did the same thing with EVTs. This is actually done in the earlier days before we have optimized the whole workflow. As you can see, the AI actually divide that into many more groups.

But it turns out they are debris, they are dead cells, or some cells, they actually went through different processes. Basically, for example, we call out of focus. That means it no longer gives you a very clear images. Okay. Excluding all that, we actually identify eight subgroups for EVTs and we haven't actually, so each one, but here just show a few examples. These are the cells. By now, we actually form a vision or a dream. That is, is it possible that we take the action cells from any tissues going through the AI network with our hardware, the ability to create label free images.

Then we isolate each group and then perform differential gene expression analysis, or proteomic analysis, or metabolic analysis. After that, we can relate the gene expression, the proteomic metabolic profile, back to the imaging features. If we can close the loop, and then after training, We hope that we can have the cells going through the system without actually doing a deeper molecular analysis, we could actually predict and generate the proteomic profile, genomic profile, and metabolic profile. That's something we do actually in the path of development, but this is something we hope to achieve.

In that case, we could actually create, at even the single cell level using AI to generate the multi-omic profile purely from imaging. This is, of course, something and also we could hopefully predict the fate of the cell and behavior of the cells to actually have a different time point, and then the last few minutes, I'd like to switch gear. How good AI can do depends on how good your original image can be. The image right now, the AI has to actually identify since we only see the 2D image for the same cell.

If you look at the front, from the side, from the back, they could the AI have to actually identify the electric from the same cells. Just like a person, the side profile and the back profile and the front profile could be actually different. If we could do the cell tomography, 3D tomography, just like we do a CT scan of every single cell, that would be very nice. There's a tool to do that we call confocal microscope on the left, but it takes actually 10 seconds to do one cell.

That's certainly too low speed. We actually increase the speed by 10,000 times, we are able to do 1,000 3D images per cell, and that is good. But then the question is, how do we know this image are real things for the cells? Nobody knows the image, okay, so we don't have a ground choose. The way we want to do it is we turn our system into a model. It's almost a concept of digital twins. We are actually using the Z mac simulation program to simulate the behavior of the system.

On the left is the simulated object. We put randomly the scattering center in the 3D space. On the right is what the digital twins generated based on our system, and then you can see there's a one to one correspondence between the generated image and the real input image, except there's some spatial resolution degradation limited by the optics. With that confirmation, then what we can do is we can actually run through the real cells, and then you can see the 3D rendering. But then if you adjust the threshold, you start to see the structure, the three dimensional inner structure of the cell, the scattering center.

With that, we actually start to analyze, in this case, we analyze the radiation damage of the cells. The middle one show a lot of red spot that is actually the double strand broken DNA by the gamma ray radiation. It's actually Gamma H2AX. The degree of the radiation damage of the glioblastoma brain tumor cell is actually by the foci count. How many sections? And here shows the center picture, it shows the fluorescence and the false eye count, and you see absolutely no correlation. In other words, you cannot just detect integrated intensity and then to determine the foci.

But with a 3D system, you could actually not only sees each broken section of the DNA, you can also index them and do an analysis. This is my last one. I realize the time is up. I'd like to acknowledge my former students and my current students. They all graduate. The system I talked about went through three generations of development. The currently, we have the last three person on the second row, Ziqi Zhang, Zhaoyu Lai, and Dr. Rahman who just joined the group. Have been actually working on the project with Louise.

Louise Laurent: I wanted to just thank the folks in my group. This is my group. The red circles are the folks who have been working on this project, especially Omar, who's there on the front in the what do you call it plaid shirt, who was the one who cultured all of the cells and ran around campus.

Speaker 1: Great. We're going to go ahead and wrap tis up because we want to give everybody a time for a break so let's thank our speakers again.