

horvath.peter.2_10_22

Life beyond the pixels: single-cell analysis

Horváth Péter

Tézisfüzet

**Szintetikus és Rendszerbiológiai Egység
Biokémiai Intézet
Szegedi Biológiai Kutatóközpont
Eötvös Loránd Kutatási Hálózat**

Szeged

2022

Bevezetés

A biológia jelenlegi ismeretanyagának nagy része, beleértve a sejhálózatok és a jelátvitel számos modelljét, populációs szintű, átlagolt méréseken alapul (Altschuler and Wu, 2010). A nagy számú sejt együttes viselkedését átlagoló mérések azonban félrevezető következtetésekhez vezethetnek: elfedik a ritka, de a teljes hálózat szempontjából kritikus alpopulációk jelenlétét (Pelkmans, 2012). Ma már jól ismert, hogy a heterogenitás még egyes kis alcsoportokon belül is fontos következményekkel járhat a populáció egészére nézve. A genetikai változatosság például döntő szerepet játszik a gyógyszerrezisztenciában és a tumorok túlélésében (Heppner, 1984).

Még azonos genetikai összetétellel rendelkező sejtcsoportokban is jelentős eltérések figyelhetők meg az egyes sejtek között a fenotípusok szintjén, a gének kifejeződésének egyedi mintázata miatt (Tay *et al.*, 2010). A heterogén biológiai rendszerek jobb megértése érdekében ezért egyre inkább az egyedi sejtekre, úgynevezett "egysejtekre" alapuló molekuláris és morfológiai elemzési módszerekre támaszkodunk (Strack, 2022). Ezeknek az egyedi sejteknek a gyűjtése, valamint a további molekuláris vizsgálatok céljából történő elkülönítése még mindig kihívást jelentenek, ahogy az is, hogy az ilyen egysejt-vizsgálatok eredményét hogyan alkalmazzuk a biológiai rendszerek viselkedését, szerkezetét és működését leíró modellek megalkotására. (Giladi and Amit, 2017).

Kutatásom célja az egyedi sejt alapú mikroszkópia, a morfológiai és fenotípusos elemzési technikák, valamint az egysejt-izolálási és molekuláris elemzési módszerek fejlesztése, ezen túlmenően a kiválasztott sejtek gyors, megbízható és sokoldalú jellemzése még a legkomplexebb sejtes mikrokörnyezetben is, mint például az emberi agykéreg, a gyermekkori agydaganatok 3D-s szferoidjai vagy a patológiai szöveti metszetek.

1. Tézis: Technikák egyedi sejtek mikroszkópiai képanalízisére

A **mikroszkópos képelemzés** során az a célunk, hogy a digitális képeket olyan kvantitatív mérésekké alakítsuk át, amelyek például képesek leírni a kísérletben szereplő minden egyes sejt állapotát. Ehhez több egymástól független módszert szükséges alkalmaznunk és fejlesztenünk.

Minden mikroszkóp által készített kép **inhomogén megvilágítást** mutat, főként a fényforrás vagy optikai útvonal sajátosságából adódóan, ami árnyékolást eredményez a kép széléin.

Ezt a hatást szemmel gyakran alábecsülik, az intenzitások azonban általában 10-30%-kal eltérnek a kép egyes részein, ami rontja például a szegmentálás pontosságát, az intenzitásmérést pedig megbízhatatlanná teszi. A megvilágítás digitális korrekciója egy olyan folyamat, amelynek segítségével a torzított képből visszaállítható a valódi kép (Smith *et al.*, 2015). Erre három fő megközelítés létezik; (1) A prospektív módszerek referencia képekből (például sötét és világos képekből) származtatják a korrekciós függvényeket. Ez a megközelítés gondos kalibrációt igényel a kép felvételekor (Singh *et al.*, 2014). (2) Az egyedi képekre alkalmazott retrospektív módszerek minden egyes felvételre külön-külön számítják ki a korrekciós modellt (Babaloukas *et al.*, 2011). Az eredmény azonban képről képre változhat, és így megváltoztathatja a relatív intenzitást; több mozaikdarabkából álló felvétel esetén egymással nem összehasonlítható részeket kaphatunk. (3) A legtöbb laboratórium által választott retrospektív, több képet alkalmazó módszerek a kísérlet során felvett képek felhasználásával utólag határozzák meg a korrekciós függvényt. Ezek gyakran energia-minimalizálási modelleken alapulnak (Peng *et al.*, 2017). Az általunk bemutatott CIDRE módszer (Smith *et al.*, 2015) is energiainimalizáció segítségével képes már nagyon kis számú kép készítésekor is felülmúlni mind a digitális mind pedig a hardveres kalibrációs technológiákon alapuló módszereket. Ez különösen fontos a nagy áteresztőképességű kvantitatív mérések esetében.

A digitális mikroszkópiában jellemzően minden egyes sejtet külön-külön azonosítanak és értékelnek ki; vagyis a pixeleket úgy csoportosítják, hogy megkülönböztessék a sejtet a többi sejtől és a háttértől. Ezt a folyamatot "**szegmentálásnak**" nevezzük, amelyre két fő módszer létezik. (1) A modell alapú megközelítés esetében választunk egy algoritmust az elkülöníteni kívánt alakzatok várható tulajdonságai (például méretük és alakjuk) alapján, és a szegmentálási eredmények vizuális vizsgálata alapján finomhangoljuk a paramétereiket. Ez a *priori* tudást (azaz "modellt") igényel (Molnar *et al.*, 2016). A modell alapú megközelítések jellemzően hisztogram-alapú módszereket alkalmaznak, például küszöbölést vagy éldetekciót, illetve watershed ("vízválasztó") transzformációt. (2) A gépi tanuláson alapuló megközelítések esetében egy modellt tanítunk a megfelelő szegmentálási megoldás megtalálására úgy, hogy előre felcímkézett (annotált) adatokat mutatunk neki, amelyek meghatározzák, hogy a kép mely képpontjai tartoznak a különböző osztályokhoz (Hollandi *et al.*, 2020). A szegmentálást végül a betanított modell új képekre történő alkalmazásával végezzük, melyek a pixeleket ennek megfelelően osztályozzák. NucleAIzer nevű módszerünk mélytanuló hálózatok kombinálásával a Data Science Bowl 2018 (DSB2018) világverseny legmagasabb pontszámát érte el (Hollandi *et al.*, 2020).

A (Caicedo *et al.*, 2017) által készített áttekintésben betekintést nyújtottunk a sejtek profilozására használt kép- és adatelemzési stratégiákba.

2. Tézis: Egysejt fenotipizálás

A tudomány eddigi egyik legnagyobb eredménye az emberi genom teljes szekvenálása (Venter, 2001). Ma sokan úgy vélik, köztük a Human Cell Atlas konzorcium kutatói is, hogy a biológia következő nagy kihívása a fenomika, azaz a fenotípusok azon halmazának számszerűsítése, amely egy szervezetet teljes mértékben jellemez (Houle, Govindaraju and Omholt, 2010). A fenotípus egy szervezet megfigyelhető jellemzőit jelenti, mint összességet, beleértve a morfológiát, a biokémiai tulajdonságokat, a viselkedést stb.. Az információban gazdag fenotípusos adatok gyűjtése és elemzése révén reméljük, hogy jobban megérthetjük, hogyan okoznak változást a szervezetekben vagy azok viselkedésében genetikai és környezeti tényezők, és képessé válunk olyan fontos jelenségek jobb előrejelzésére, mint a fitness, a szaporodás, bizonyos betegségek, a karcinogenezis, a különböző rezisztenciák kialakulása vagy a mortalitás. A genomszekvenálással ellentétben a fenom teljes megértése a jelenlegi technológiákkal egyelőre lehetetlen. Az erre irányuló erőfeszítések során kritikus fontosságú, hogy intelligens döntéseket hozzunk arról, hogy mit mérjünk, és milyen fenomikai eszközöket használjunk.

A képkalkotás gyors és rugalmas módszer a fenomika tanulmányozására. A térbeli és időbeli információk nagy pontossággal és rendkívül széles skálán rögzíthetők. A mikroszkópia implicit módon mutatja meg a sejt morfológiai jellemzőit, és az olyan jelölési technológiák, mint a fluoreszcens címkék lehetővé teszik a sejtalkotók, fehérjék és más molekulák helyének és alakjának pontos meghatározását. A képek alacsony költséggel és gyorsan készíthetők, ami lehetővé teszi nagyléptékű szűrési kísérletek tervezését és kivitelezését. A mikroszkópia, az automatizálás és a számítástechnika terén a közelmúltban bekövetkezett fejlődés drámaian megnövelte a fenotípusos információban gazdag képek készítésének lehetőségét. Emellett a képek ma már nagyságrendekkel gyorsabban készíthetők, mint ahogyan azokat kizárólag emberi erővel vizsgálni lehetne. Következésképpen fenotípusos képelemzési technikákra támaszkodunk, azaz olyan számítási módszerekre, amelyek a nyers képi adatokat kvantitatív fenotípusos adatokká alakítják. E terület átfogó áttekintését (Smith *et al.*, 2018) munkájában mutattuk be .

Csoportom számos olyan eszközt és szoftvercsomagot fejlesztett ki, amelyek intelligensen segítenek meghatározni az egyes sejtek fenotípusát. Mindegyik kutatási projekt a következő két kérdésre keresi a választ:

'Teljesen vagy legalább részben feltártam-e az adataimat?'

'A lehető legpontosabb-e az elemzésem?'

A modern számítástechnikai módszerek ma már elképzelhetetlen mennyiségű adatot képesek rövid idő alatt előállítani. Sokszor ez több százezer képet, több milliárd egyedi sejtet és

gyakran százas nagyságrendű egyedi jellemzőt tartalmaz minden egyes sejtről. Ekkora adatmennyiség mellett a fenti két fő kérdés megválaszolása már nem egyértelmű.

A 2010-es évek elején azzal a céllal vezettük be az **Advanced Cell Classifier** (ACC, elérhető a www.cellclassifier.org címen) nevű szoftvert (Horvath *et al.*, 2011), hogy az egysejtes fenotípezést végző szakértők számára könnyen hozzáférhetővé tegyük a gépi tanulási módszereket. Ezt a programot később kibővítettük (ACC2.0) úgy, hogy a fenti kérdésekre intelligens fenotípuskereső és aktív tanulási módszerek alkalmazásával adjon választ (Piccinini *et al.*, 2017). Az elmúlt évtizedben az ACC különböző változatait új gyógyszerek és gének felfedezésében, a biológiai alapkutatással kapcsolatos kérdések megválaszolásában, valamint az egyénre szabott terápiák kifejlesztésében is hasznosították. Rangos publikációk jelentek meg a Cell, a Science és a Nature folyóiratok hasábjain (a hivatkozási listát lásd az ACC weboldalon).

Jelentős mérföldkőnek számít, hogy kutatócsoportjaim lefektették annak az elméleti keretnek az alapjait, amely a sejtek **mikrokörnyezetének fontosságát** hangsúlyozza, azaz a sejtek fenotípusai pontosabban határozhatók meg, ha ismerjük azok környezetét a szöveten belül. Már a korai módszerünk, amelyben csak egyszerű tulajdonságokat és szomszédsági meghatározásokat vettünk figyelembe, jelentős javulást eredményezett (Toth *et al.*, 2018). Ennek a technikának a kiterjesztése, amely a halszem-transzformációt mélytanulással kombinálta, jelentősen javította az egysejtes fenotípusok meghatározását, mind szövet-, mind sejt kultúrák mintákban (Toth *et al.* 2022 in press).

A sejt szinte folyamatosan **dinamikus változáson** megy keresztül, ami a fenotípusának megváltozásával is jár. Dolgozatomban részletesen bemutatom az általunk javasolt megoldást az egyes sejtek folytonos variációjának absztraktabb számszerűsítésére mesterséges intelligencia felhasználásával. Ehhez a feladathoz bevezettük a **regressziós sík koncepciót** is (Szkalisity *et al.*, 2021).

3. Tézis: Egysejt izolálás

Számos módszer használható a sejtek összegyűjtésére bizonyos jellemzők alapján. Ezek közé tartozik a fluoreszcencia-aktivált sejtválogatás (FACS), az immunmágneses sejtválogatás, a mikrofluidika és a limitáló hígítás. Ezek a gyűjtési technikák azonban megbontják a szövet egységét, elválasztják a sejteket a mikrokörnyezetüktől, valamint nem tudják a sejteket a mintán belüli elhelyezkedésük vagy morfológiai profiljuk alapján kiválogatni. Ezzel szemben a mikromanipuláció (pl. patch clamping), a lézeres mikrodisszekció (Espina *et al.*, 2006), a képkövető tömegspektrometria (IMS) (Bodzon-Kulakowska and Suder, 2016) vagy a Raman-spektroszkópia (Schie and Huser, 2013) olyan mikroszkópián alapuló

módszerek, amelyek segítségével közvetlenül elemezhetjük az egyes sejteket akár a szövetmintákon belül is.

Kutatócsoportjaim a szubmikron felbontású képalkotást, az egysejtes fenotipizálást és a mesterséges intelligencia által támogatott izolálást kombinálták elektrofiziológiai, szekvenálási és ultraérzékeny proteomikai munkafolyamatokkal. A legnagyobb kihívás az egysejt határok és fenotípusos sejtsztyályok pontos meghatározása. Szoftvereszközeink képesek egyszerre alkalmazni a klasszikus, a lézeres mikrodisszekciós és a patch clamp mikroszkópiát. Ezeket aztán kombinálhatjuk a sejtkultúrák vagy archivált FFPE (formalinban rögzített és paraffinba ágyazott) biobanki szövetek adatgazdag képalkotásával, mély tanuláson alapuló sejt szegmentálással és a sejttípusok és állapotok gépi tanuláson alapuló azonosításával. A biológiai és/vagy klinikai szempontból érdekes sejt vagy szubcelluláris objektumokat mesterséges intelligencia választja ki, melyekről ezután automatizált omikai profil készül. Az így előállított adatok segítségével olyan RNS és fehérje mintázatok felfedezésére nyílik mód, amelyek molekuláris betekintést nyújtanak a genom, a proteom vagy más omikák fenotípus szintű változatosságába, mindeközben megőrizve a teljes környezeti információt.

A közelmúltban az egysejtes izolációs módszerek négy családját mutattuk be, nevezetesen (1) a számítógéppel segített mikroszkópiás képalkotást és izolációt (**CAMI**) (Brasko *et al.*, 2018), amely egy **gépi tanuláson alapuló egysejtes izolációs** technika, és lézeres mikrodisszekciót alkalmaz 2D mikrokörnyezetben; (2) az AutoPatcher rendszert, amely teljesen automatikusan képes a sejteket **3D élő sejt** környezetben megtalálni egy patch clamp technika és mély tanulás kombinációjával (Koos *et al.*, 2021); (3) a **Deep Visual Proteomics módszert**, amely a CAMI rendszerből származik, és ultraérzékeny proteomikára alkalmazható (Mund *et al.*, 2022); és (4) a Mito-Raman technikát, amely **intelligens mikroszkópiát és Raman-spektroszkópiát** használ a mitotikus sejtek automatikus profilalkotására (Voros *et al.*, előkészületben). Ezek a módszerek az általunk eddig kifejlesztett képelemzési és mélytanulási technikákra épülnek. Célunk, hogy újabb módszerek beépítésével ezeket továbbfejlesszük.

Konklúzió és kitekintés

Dolgozatomban bemutattam, hogyan lehet morfológiailag és molekulárisan is profilozni az egyes sejteket, még komplex térbeli környezet esetén is. Képkorrektációs és szegmentálási módszereket dolgoztam ki a mintákban lévő sejtek pontos meghatározására, valamint fenotipizáló algoritmusokat fejlesztettem ki a sejttípusok automatikus azonosítására és

horvath.peter.2_10_22

osztályozására. Az egysejtes manipulációs technikákkal képesek vagyunk a kívánt sejtet elkülöníteni eredeti környezetéből a részletes molekuláris jellemzés érdekében.

Ezen technológiákkal úgy vélem, hogy megnyitottuk az ajtót a pontosabb egysejtes felfedezések előtt. A bemutatott módszerek legtöbbje általános labor környezetben is megvalósítható, és lehetővé teszik a tudósok számára, hogy elvégezzék az őket érdeklő méréseket. Számos érdekes, a bemutatott munkafolyamatokon alapuló, kutatást lehet előrevetíteni.

A dolgozatban bemutatott előrelépések ellenére úgy gondolom, hogy ez még csak a kezdet, és még sok a tennivaló. Kutatócsoportjaim tervei közé tartozik a jövőben az itt leírt területeken való további előrelépés, valamint a kifejlesztett technológiák felhasználása mind az alapkutatásban (lásd alább a Human Mitosis Atlas-t), mind az egysejtes rákterápiákban (hopeAI). Itt megosztanám az olvasóval a véleményemet arról, hogyan képzelem el e területek fejlődését.

Valószínűleg a képelemzés profitált a legtöbbet a mélytanulás térnyeréséből, és ez minden bizonnyal csak a kezdet. A hardverek exponenciális fejlődésével és a kvantum számítástechnikával, mely már az ajtóban kopogtat, jó okunk van azt feltételezni, hogy sokkal nagyobb mennyiségű képi adatot leszünk képesek digitálisan feldolgozni, és szisztematikusabb kérdéseket feltenni, például a rákos megbetegedések jellemző struktúráira és mintázataira a populációkban egyetlen sejt szintjén. Ennek érdekében folytatjuk a mélytanulási eszközök fejlesztését, amelyek egyre jobb szegmentálást és a képi információ egyre jobb globális megértését eredményezik. Egy másik terület, amely a közeljövőben valószínűleg tovább fog fejlődni, az a 3D-s képképzés és feldolgozás, mivel a szferoid, organoid és más szövetszerkezetes 3D-s modellek jobban közelítik a sejtek viselkedésének fiziológiáját, mint a "sík" biológia. Valójában már megtettük az első lépéseket e cél felé is, és a jövőben több erőforrást fogunk fordítani a 3D képelemzési modellek szabványosítására, felgyorsítására és javítására.

A bemutatott eszközökkel a kezünkben már szintén megtettük az első lépéseket egy nagyszabású kezdeményezés, a Human Cancer Mitosis Atlas (MITO-Omics) felé. Ez átfogóan, közel végtelen felbontásban fogja leírni a sejtek mitózis során bekövetkező molekuláris és morfológiai változásait. Tervünk, hogy a jelenlegi adatainkat kiterjesszük az összes emberi daganattípusra. A MITO-Omics nagyszerű forrás lesz a gyógyszerfejlesztés és a klinikai alapkutatások számára a rákos megbetegedések jobb megértése és a rák elleni küzdelem érdekében.

Végül személyes célom, hogy az egysejtes módszereinket a személyre szabott rákterápiában is bevezessük. Ennek kapcsán szeretném a rákos megbetegedéseket egysejtes szinten profilozni, hogy a feltárt genomikai és proteomikai elváltozások alapján precíziós terápiák javaslatára legyen lehetőségünk. Ez minden bizonnyal az akadémiai és az ipari szféra közös vállalkozása lesz. Az első lépéseket már megtettük ezen cél felé, és az első klinikai sikertörténet már meg is született (nem publikált).

Irodalomjegyzék

- Altschuler, S.J. and Wu, L.F. (2010) 'Cellular heterogeneity: do differences make a difference?', *Cell*, 141(4), pp. 559–563.
- Babaloukas, G. *et al.* (2011) 'Evaluation of three methods for retrospective correction of vignetting on medical microscopy images utilizing two open source software tools', *Journal of microscopy*, 244(3), pp. 320–324.
- Bodzon-Kulakowska, A. and Suder, P. (2016) 'Imaging mass spectrometry: Instrumentation, applications, and combination with other visualization techniques', *Mass spectrometry reviews*, 35(1), pp. 147–169.
- Brasko, C. *et al.* (2018) 'Intelligent image-based in situ single-cell isolation', *Nature communications*, 9(1), p. 226.
- Caicedo, J.C. *et al.* (2017) 'Data-analysis strategies for image-based cell profiling', *Nature methods*, 14(9), pp. 849–863.
- Espina, V. *et al.* (2006) 'Laser capture microdissection', *Methods in molecular biology*, 319, pp. 213–229.
- Giladi, A. and Amit, I. (2017) *Immunology, one cell at a time*, Nature Publishing Group UK. Available at: <https://doi.org/10.1038/547027a>.
- Heppner, G.H. (1984) 'Tumor heterogeneity', *Cancer research*, 44(6), pp. 2259–2265.
- Hollandi, R. *et al.* (2020) 'nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer', *Cell systems*, 10(5), pp. 453–458.e6.
- Horvath, P. *et al.* (2011) 'Machine learning improves the precision and robustness of high-content screens: using nonlinear multiparametric methods to analyze screening results', *Journal of biomolecular screening*, 16(9), pp. 1059–1067.
- Houle, D., Govindaraju, D.R. and Omholt, S. (2010) 'Phenomics: the next challenge', *Nature reviews. Genetics*, 11(12), pp. 855–866.
- Koos, K. *et al.* (2021) 'Automatic deep learning-driven label-free image-guided patch clamp system', *Nature communications*, 12(1), p. 936.
- Molnar, C. *et al.* (2016) 'Accurate Morphology Preserving Segmentation of Overlapping Cells based on Active Contours', *Scientific reports*, 6, p. 32412.

- Mund, A. *et al.* (2022) 'Deep Visual Proteomics defines single-cell identity and heterogeneity', *Nature biotechnology*, 40(8), pp. 1231–1240.
- Pelkmans, L. (2012) 'Cell Biology. Using cell-to-cell variability--a new era in molecular biology', *Science*, 336(6080), pp. 425–426.
- Peng, T. *et al.* (2017) 'A BaSiC tool for background and shading correction of optical microscopy images', *Nature communications*, 8, p. 14836.
- Piccinini, F. *et al.* (2017) 'Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data', *Cell systems*, 4(6), pp. 651–655.e5.
- Schie, I.W. and Huser, T. (2013) 'Methods and applications of Raman microspectroscopy to single-cell analysis', *Applied spectroscopy*, 67(8), pp. 813–828.
- Singh, S. *et al.* (2014) 'Pipeline for illumination correction of images for high-throughput microscopy', *Journal of microscopy*, 256(3), pp. 231–236.
- Smith, K. *et al.* (2015) 'CIDRE: an illumination-correction method for optical microscopy', *Nature methods*, 12(5), pp. 404–406.
- Smith, K. *et al.* (2018) 'Phenotypic Image Analysis Software Tools for Exploring and Understanding Big Image Data from Cell-Based Assays', *Cell systems*, 6(6), pp. 636–653.
- Strack, R. (2022) 'Spatial proteomics with subcellular resolution', *Nature methods*, p. 780.
- Szkalicity, A. *et al.* (2021) 'Regression plane concept for analysing continuous cellular processes with machine learning', *Nature communications*, 12(1), p. 2532.
- Tay, S. *et al.* (2010) 'Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing', *Nature*, 466(7303), pp. 267–271.
- Toth, T. *et al.* (2018) 'Environmental properties of cells improve machine learning-based phenotype recognition accuracy', *Scientific reports*, 8(1), p. 10085.
- Venter, J.C. (2001) *The Sequence of the Human Genome: Figure 1.*