



Universiteit
Leiden
The Netherlands

Data analysis for mass spectrometry imaging : methods and applications

Abdelmoula Walid, Mohamed

Citation

Abdelmoula Walid, M. (2017, January 18). *Data analysis for mass spectrometry imaging : methods and applications*. *ASCI dissertation series*. Retrieved from <https://hdl.handle.net/1887/45501>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/45501>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/45501> holds various files of this Leiden University dissertation

Author: Abdelmoula, Walid M.

Title: Data analysis for mass spectrometry imaging : methods and applications

Issue Date: 2017-01-18

SAMENVATTING EN CONCLUSIES

In dit proefschrift hebben we een aantal automatische methoden voor multimodale data registratie ontwikkeld, vooral tussen massaspectrometrie imaging, optische microscopie en de Allen Brain Atlas. We hebben het belang van deze methoden aangetoond door het uitvoeren van grootschalige preklinische biomarker onderzoek op het gebied neurologische aandoeningen. We hebben ook een data-gestuurde benadering voorgesteld om tumorweefsel van patiënten in te delen in moleculair verschillende tumor subpopulaties om vervolgens automatisch die tumor subpopulaties te identificeren die aan de patiënten uitkomst zijn gelinkt. In de volgende paragrafen geven we een beknopt overzicht van het onderzoek, discussie, en toekomst perspectief van de ontwikkelingen.

TECHNISCHE ONTWIKKELINGEN

CO-REGISTRATIE TUSSEN MSI DATA EN DE ALLEN BRAIN ATLAS

In **Hoofdstuk 2** hebben wij een beeldbewerking pijplijn voorgesteld om automatisch MALDI MSI datasets uit verschillende muizenhersenen aan de Allen Brain Atlas (ABA) te koppelen. De voorgestelde pijplijn bestaat uit drie belangrijke stappen: (1) het voorbewerken van histologische beelden, (2) de automatische selectie van de doel gebieden in de ABA, en (3) de niet-rigide registratie. De registratie resultaten werden kwantitatief beoordeeld en de complete registratie fout ($30\ \mu\text{m}$) was beneden de pixel resolutie van de gebruikte MSI data ($100\ \mu\text{m}$). Deze automatische MSI-ABA registratie pijplijn maakt het gemakkelijk om verschillen tussen dieren te onderzoeken, de hoeveelheid van biomoleculen per gebied te kwantificeren en het correleren van de MSI gegevens aan de genexpressie gegevens uit de ABA.

Handmatige annotatie door visuele inspectie van aangekleurd weefsels blijft de standaard procedure bij histologie/anatomie die wordt gebruikt om MSI data-analyse te sturen. MSI is meestal geïmplementeerd in massaspectrometrie laboratoria en dus een dergelijke annotatie is de belangrijkste beperking voor deze biomedische toepassingen, vooral in het geval wanneer de preklinische studie de analyse van een groot aantal monsters vereist. Het ontwikkelen van automatische registratie en annotatie gereedschappen om de vele monsters te kunnen verwerken zal niet alleen de doorvoer van de data-analyse verhogen, maar ook de vergelijking van gegevens van MSI uit grote cohorten vergemakkelijken; i) door het extraheren van gegevens uit bepaalde gebieden van de hersenen; ii) door vergelijking van MSI verdelingen tussen verschillende dieren in dezelfde coördinaten systeem; en iii) doordat de correlatie tussen de grote MSI datasets met bijvoorbeeld de anatomische of genexpressie gegevens in de ABA beschikbaar is.

De selectie van het beste referentie beeld uit de ABA, dat wil zeggen de stereotactische locatie van de experimentele weefselsectie in de referentie weefselatlas, is van het grootste belang voor het succes van het registratieproces. De ABA bevat 132 coronale en 21 sagittale secties gelijkmatig verdeeld op een afstand van $100\ \mu\text{m}$ en $200\ \mu\text{m}$ respectievelijk, verspreid over de gehele hersenen van muizen. In het preklinische onderzoek naar corticale verspreiding van depressie beperkt het zoekgebied in de ABA zich tussen de bregma posities $-1,22$ en $-2,80\ \text{mm}$, oftewel 14 referentie coronale secties. Binnen dit gebied hebben we vastgesteld dat de laterale omvang van de hippocampus lineair gecorreleerd met de positie van het referentiepunt en kan worden gebruikt voor de nauwkeurig en automatische bepaling van de beste match van de experimentele

weefselcoupe met de coupe uit de ABA. Soortgelijke benaderingen kunnen worden gebruikt voor experimentele weefselcoupes die zijn gesneden op verschillende locaties, maar dan moet er een andere metrische kenmerk worden gebruikt om de beste match in de ABA te bepalen. De huidige methode corrigeert niet voor significante afwijkingen in de hoeken van het snijvlak ten opzichte van de ABA; grote verschillen kunnen de mogelijkheid om biologische conclusies te trekken uit de ABA bemoeilijken.

Het koppelen van de MSI gegevens aan de ABA werd uitgevoerd door het registreren van de experimentele histologie met het referentiebeeld histologie beeld van de ABA en dan het toepassen van dezelfde transformatie matrix op de MSI-data. Daarom is een voorwaarde voor deze MSI-ABA registratie pijplijn, dat de MSI gegevens al moeten zijn geregistreerd aan het histologische beeld. Op dit moment is deze aanpak om MALDI-MSI data verkregen beperkt omdat alleen data die met behulp van Bruker Daltonics machines kan worden gebruikt omdat deze commerciële verkoper deze aanpak heeft beschermd. De FlexImaging software van Bruker Daltonics biedt semiautomatische registratie software op basis markers. Er was geen generieke oplossing beschikbaar voor MSI datasets opgenomen met behulp van een MSI massaspectrometer. Daarom hebben we deze beperking aangepakt in **Hoofdstuk 3** om automatisch MSI gegevens te registreren aan het bijbehorende histologische beeld en hebben succesvol laten zien in **Hoofdstuk 7** dat het mogelijk is om MSI datasets van verschillende massa spectrometers te registreren aan hun histologie afbeeldingen en de ABA.

De voorgestelde pijplijn is gebaseerd op het gebruik van de ABA als referentie atlas, het is echter ook mogelijk om andere referentie muis hersenen atlassen (zoals Paxinos atlas) te gebruiken. ABA werd gekozen omdat het unieke en publiekelijk beschikbare data bevat van multimodale beeldvorming van de hersenen van muizen zoals magnetische resonantie beelden, hoge resolutie histologische afbeeldingen, hoge resolutie neuronale connectiviteit beelden, en gen expressies van meer dan 20.000 genen. De voorgestelde pijplijn kan toekomstig neuro-proteomics onderzoek mogelijk te maken met behulp van MSI. Dit kan een beter inzicht in de pathofysiologie van neurologische aandoeningen geven en leiden tot de ontwikkeling van nieuwe therapieën.

Deze pijplijn maakt het mogelijk om grootschalige preklinische MSI onderzoek van neurologische aandoeningen in de hersenen van muizen te doen. In **Hoofdstuk 6** hebben we de bruikbaarheid van de MSI-ABA aangetoond met een pijplijn voor het analyseren van de biomoleculaire veranderingen die gepaard gaan met het verspreiden van corticale depressie in een muismodel van migraine. Op dezelfde wijze kan deze pijplijn worden toegepast om moleculaire biomarkers in muismodellen van andere neurologische aandoeningen zoals de ziekte van Alzheimer, Beroerte en de ziekte van Parkinson te ontdekken.

CO-REGISTRATIE TUSSEN MSI DATA EN HISTOLOGIE

Het is bewezen dat de integratie van de biomoleculaire gegevens verkregen met MSI met de anatomische structuur door histologie essentieel is voor de klinische en farmacologische toepassing. In **Hoofdstuk 3** presenteerden we een automatische co-registratie methode om MSI dataset aan de histologische beelden van weefsel koppelen. We hebben ook aangetoond dat de voorgestelde methode algemeen toepasbaar is voor de data van verschillende organen, verschillende massaspectrometers en verschillende io-

nisatie methoden. De registratie kwaliteit is gekwantificeerd en de algemene registratie nauwkeurigheid was voldoende met een fout minder dan de grootte van een enkele MSI pixel.

De t-SNE representatie speelt een belangrijke rol in het MSI-histologie registratie pijplijn door de reductie van de spatiomolecular verdeling over het weefsel, die duidelijke overeenkomsten heeft met histologie van de weefselcoupe. De computationele complexiteit van de t-SNE is een cruciale factor in dit proces. De oorspronkelijke implementatie van de t-SNE heeft een kwadratische computationele complexiteit en bijbehorend geheugen gebruikt en daardoor een beperkte capaciteit van 10.000 datapunten. Onlangs zijn verdere verbeteringen gemaakt om de t-SNE te versnellen op basis van de Barnes-Hut implementatie (BH-SNE), die schaalst logarithmisch voor de berekeningen en lineair voor het geheugengebruik, waardoor t-SNE analyse van MSI gegevensverzamelingen veel meer de praktijk kan worden toegepast. De huidige ontwikkelingen in de MSI-technologie die 3D MSI gegevens kunnen verzamelen [209] is een computationeel uitdaging, zelfs voor de Barnes-Hut implementatie. Echter de 3D MSI gegevens bestaan uit een stapel van opeenvolgende coupes van afzonderlijke MSI datasets - dus de Barnes-Hut implementatie kan worden toegepast om elke afzonderlijke weefselcoupe aan de MSI gegevens set te registreren, waarna aan de hand van de histologische beelden de MSI datasets van de opeenvolgende beelden kunnen worden geregistreerd.

HISTOLOGIE GEDREVEN MASSA SPECTROMETRIE

Hogere spatiele/massa resolutie leidt tot een toegenomen omvang van data en langere data-acquisitie tijden. Voor klinische toepassingen, waarbij grote hoeveelheden patiëntweefsel moeten worden geanalyseerd, vormt dit een uitdaging om de hoeveelheid data en de acquisitie tijd beheersbaar te houden. Daarom is in **Hoofdstuk 4** een geautomatiseerde beeldregistratie pijplijn ontwikkeld voor het registreren een geannoteerde histologische beeld van een aangrenzend weefselsectie aan een lagere resolutie weefselbeeld van de MSI (matrix geccoat) weefselsectie. Vervolgens worden de annotatie contour overgezet naar het lage-resolutie beeld waardoor alleen die gebieden worden geanalyseerd op de MSI. Deze-histologie gestuurde MSI (HG-MSI) aanpak garandeert dat de MSI gegevens uitsluitend verkregen wordt uit weefsel regio's met een vergelijkbare histologische samenstelling. We hebben aangetoond dat een reductie van 80% van de hoeveelheid data en acquisitie tijd mogelijk is, waardoor het mogelijk wordt om hoge resolutie (massa of ruimtelijke) toe te passen in klinisch onderzoek.

De begrenzing van ROI voor de HG-MSI aanpak hangt vooral af van het de mate van detail in de histologische beelden. Terwijl een bepaald gebied als histologisch homogeen wordt weergegeven, kan de moleculaire samenstelling heterogeen zijn. De behoefte aan histologische specificatie is afhankelijk van de toepassing, b.v. de analyse van de tumor interfaces, of het onderzoek naar de overeenkomsten / verschillen van goed versus ongedifferentieerde tumorcellen. Het is dus cruciaal om in het begin te specificeren waarvan histologische vergelijkingen worden/kunnen gemaakt. De tool is het best geschikt voor analyses waarin de langzame scansnelheid (hoge massa resolutie) of groot aantal pixels (hoge ruimtelijke resolutie) het onpraktisch maakt om het toe te passen op complete weefselsectie van patiënten.

We hebben laten zien dat deze HG-MSI benadering werkt voor de analyse van de proteolytische peptiden van specifieke histopathologische regio's in tumorweefsels met behulp van een ultra hoge massa resolutie Fourier transform ion cyclotron resonantie spectrometrie van Bruker Daltonics. Echter, de tool is onafhankelijk van de fabrikant en kan worden toegepast op alle MSI methodologieën (Bijvoorbeeld DESI, SIMS, enz.); het enige vereiste is dat een optisch beeld van de weefselsectie wordt gebruikt om de meetgebieden definiëren. Evenzo de HG-MSI software kan worden gebruikt om te focussen op hoge spatiele resolutie evenals op hoge massa resolutie analyse, en kan worden gebruikt voor alle massa klassen en alle soorten weefsel.

INTRA-TUMOR HETEROGENITEIT

Massaspectrometrie imaging heeft een groot potentieel in het onderzoek naar kanker zoals is aangetoond dat het in staat is om moleculaire verschillen te ontdekken in tumor subpopulaties in histologisch uniforme gebieden van tumorweefsel. In **Hoofdstuk 5** hebben we een automatische data gestuurde aanpak gepresenteerd, gebaseerd op de ruimtelijke verdeling in t SNE, welke in staat was om klinisch relevante tumor subpopulaties in de MALDI-MSI molecuair te onderscheiden in datasets van monsters van tumorpatiënten. We hebben statistisch aangetoond dat het mogelijk is om subpopulaties te onderscheiden van de overlevingskansen van patiënten met primaire maagkanker tumoren en lymfeklieren metastase in primaire tumoren van borstkanker. De identificatie van deze fenotypische tumor subpopulaties is van groot belang voor het begrijpen van de tumor en het verdere onderzoek betreffende de belangrijkste moleculaire veranderingen die gepaard gaan met de mogelijke tumorprogressie.

De robuustheid van de resultaten zijn gemeten met een unbiased kruisvalidatie experiment. Dit gaf de mogelijkheid om de stabiliteit van de potentiële prognostische m/z waarden die de nadelige subpopulaties karakteriseren te onderzoeken. We waren ook in staat om de pixel in te delen en de patiënten in te delen naar uitkomst, deze indeling in staat om de intra-tumor heterogeniteit te bepalen en overleving van patiënten (primaire maagkanker) te voorspellen of om de lymfeklier metastase-status (primaire borstkanker) te voorspellen.

In zowel de maag- en borstkanker cohorten was de onderliggende moleculaire en fenotypische Intra-tumor heterogeniteit niet duidelijk zichtbaar in de conventionele histologische beelden. Dit opent de nieuwe mogelijkheden voor de beoordeling van de klinische impact van Intra-tumor heterogeniteit, bijvoorbeeld via de pixel classificatie van een ruimtelijk bepaalde steekproef voor RNA-sequencing van de tumor subpopulaties met verschillende klinische fenotypen.

PRE-KLINISCHE TOEPASSINGEN

De automatische MSI-ABA registratie pijplijn zoals gepresenteerd in **Hoofdstuk 2** heeft een groot potentieel in het faciliteren van grootschalige MSI onderzoek in muizen-hersenen modellen van neurologische aandoeningen. Wij hebben dit aangetoond in **Hoofdstuk 6** door de moleculaire veranderingen te onderzoeken die geassocieerd zijn met corticale verspreiding van depressie in een migraine muismodel. Gegevens van cohorten die 3 moleculaire klassen omvatten van 32 muizen met twee verschillende

genotypen, zijn opgenomen met MALDI-MSI. Alle MSI datasets werden geregistreerd op de ABA met behulp van de ontwikkelingen gepresenteerd in **Hoofdstuk 2**. We konden daardoor gebruik maken van de anatomische annotaties in de ABA referentie atlas om de door MSI moleculaire gedetecteerde handtekeningen te vergelijken in specifieke anatomische structuren van belang (Isocortex, Thalamus, Hippocampus, en Striatum). We hebben een aantal biomoleculaire kenmerken (m/z) gevonden die significante ($p < 0.05$) anders waren in deze anatomische gebieden. Deze zijn allemaal gevonden in de transgene muis omdat deze een sterkere CSD fenotype vertoont (dus de moleculaire veranderingen zijn naar verwachting groter).

Een andere preklinische toepassing is gepresenteerd in **Hoofdstuk 7** waarin SIMS-MSI is gebruikt om de lipide-ophoping in de hersenen van een transgene muis te onderzoeken die geen multifunctioneel eiwit 2 (MFP2) heeft. MFP2 is een eiwit dat verantwoordelijk is voor de vorming van galzuren en de afbraak van pristaanzuur en de lange keten vetzuren (VLCFA). Muizen die aan MFP2 gebrek lijden, kampen met ernstige neuro motorische disfuncties en sterven vóór de leeftijd van 6 maanden. Het is dus van belang om te onderzoeken welke biomoleculen significant geassocieerd met deze pathologie en waar ze anatomisch te lokaliseren zijn. We gebruikten de ontwikkelingen die zijn beschreven in de **Hoofdstukken 2 en 3** om de SIMS-MSI datasets te registreren aan de ABA. De t-SNE registratie pijlpijn, gepresenteerd in het **Hoofdstuk 3**, werd eerst gebruikt om de SIMS-MSI gegevens te registreren aan de bijbehorende histologische beelden, om te voldoen aan de vereisten om de MSI-ABA registratie pijlpijn uit te voeren. Deze registratie pijlpijn maakte het mogelijk om het geaccumuleerde lipide in MFP2-deficiënte in transgene muizen in de 4^e ventrikel te lokaliseren. Deze precieze lokalisatie van het MSI signaal gaf een dieper inzicht in het pathologische mechanisme van de MFP2 deficiëntie.

TOEKOMSTIG WERK

In dit proefschrift hebben we een automatische registratie pijlpijn ontwikkeld om MSI data te koppelen aan de ABA. Het doel is tweeledig: 1) de koppeling MSI data van verschillende objecten aan dezelfde coördinaatruimte maakt het mogelijk om hun biomoleculaire profielen gemakkelijker met elkaar te vergelijken, 2) de anatomische annotaties die door ABA worden verschaft maken het de gebruikers van MSI gemakkelijker om de MSI data in een anatomische context te plaatsen. Toch kan dit werk verder worden uitgebreid door te kijken naar de genexpressies locaties in de ABA. De ABA toont expressies van meer dan 20000 genen in de hele hersenen van muizen. Analyse van deze genexpressiepatronen is niet triviaal en moet gebeuren met speciale computationele methoden om deze big data te verwerken. Uiteindelijk nu de integratie van MSI gegevens met deze gen uitdrukkingen haalbaar is, zou dit meer diepgaande inzichten kunnen verschaffen voor een beter begrip van de pathofysiologie van neurologische aandoeningen. Meer in het algemeen zou het de deuren kunnen openen voor veelbelovende initiatieven voor de toekomst van biologisch onderzoek.

Er is op dit moment een grote belangstelling voor 3D MSI, zodat de volledige biomoleculaire samenstelling van de weefsels kan worden onderzocht [88–90]. De huidige ontwikkelingen in de MSI-technologie, in het bijzonder de snelle MSI instrumenten, kunnen 3D MSI in veel korter tijdsbestek uit voeren. Toch de data reconstructie

en de verwerking zijn nog steeds uitdagend [90, 209]. 3D MSI data reconstructie algoritmen zouden rekening moeten houden met het grote potentiële probleem van niet-lineaire vervormingen tussen opeenvolgende weefselcoupes en nog belangrijker het definiëren van de geometrische beperkingen die nodig zijn om de oorspronkelijke vorm kenmerken te behouden en dus het vermijden van de "Banana-into-cilinder-effect" [44, 210]. Bij geautomatiseerde en nauwkeurige reconstructie 3D MSI datasets gaat het om twee afzonderlijke beeldregistratie stappen: 1) het registreren van elke individuele MSI datacube van een weefselcoupe aan het bijbehorende histologische beeld, bijvoorbeeld met behulp van de t-SNE gebaseerde geautomatiseerde registratie zoals gepresenteerd in **Hoofdstuk 3**; en 2) het gebruik de histologische beelden om ruimtelijke overeenkomsten tussen opeenvolgende coupes vast te stellen, door gebruik te maken van vergelijkbare histologie gebaseerd maatbepalingen zoals gepresenteerd in **Hoofdstuk 2** voor uitlijning met de ABA, maar nu met een extra stap van het integreren van geometrische beperkingen en deze moeten verder worden onderzocht.

De automatische registratie tussen MSI en histologie kan worden beschouwd als een basis voor MSI image fusie [[211], voor het verbeteren van de beeldkwaliteit en het verhogen van de interpreteerbaarheid van de MSI data. Onlangs hebben Van der Plas *et al* [39] een data fusie pijplijn gepresenteerd tussen MSI en histologie, maar deze vergt eerst een afstemming tussen MSI en histologie. De t-SNE gebaseerde registratie pijplijn, zoals gepresenteerd in **Hoofdstuk 3**, zou een dergelijke fusie pijplijn eenvoudiger uit te voeren maken, niet alleen op een bepaalde MSI modaliteit maar ook tussen moleculaire data van verschillende imaging massa spectrometers.