



의학석사 학위논문

포르말린 고정 조직

(Formalin-Fixed Paraffin-Embedded Tissue) 단백질체 분석을 위한 전처리 최적화 연구 Optimization of sample preparation for formalin-fixed paraffin-embedded tissue proteome analysis

울 산 대 학 교 대 학 원

의 학 과

정 황 교

포르말린 고정 조직

(Formalin-Fixed Paraffin-Embedded Tissue) 단백질체 분석을 위한 전처리 최적화 연구

지도교수 김 경 곤

이 논문을 의학석사 학위 논문으로 제출함

2022 년 2 월

울 산 대 학 교 대 학 원

의 학 과

정 황 교

정황교의 의학석사학위 논문을 인준함

심사위원	오 수 진	인
심사위원	이 희 진	인
심사위원	김 경 곤	인

울 산 대 학 교 대 학 원

2022 년 2 월

Abstract

Formalin-Fixed Paraffin-Embedded (FFPE) tissue 는 병리적 검사가 완료되어 병리과에서 보관 중이며 두께 조절을 통해 시료양을 조절하고 tumor 조직만 선별할 수 있다는 장점을 가지고 있다.

많은 장점을 가진 FFPE tissue 시료지만 제작과정에서 protein-protein crosslinking 을 형성하기 때문에 단백질 추출 및 단백질체 분석에 제한을 받게 된다. 이번 연구에서 동일한 FFPE tissue 시료를 이용하여 최적화된 단백질 추출 방법을 확인하고 147 개의 TNBC FFPE 시료에 대해 최적화된 방법으로 단백질을 추출하여 LC-MS 분석을 수행하여 치료 예후와 재발에 대한 바이오마커 발굴을 수행하였다.

2um 의 FFPE tissue slice 를 extraction buffer, AFA sonication 사용 유무, boiling 조건 등을 다르게 하여 3 가지 조건으로 선정하여 단백질을 추출한 후 단백질 정량 하였다. S-Trap digestion 방법을 이용하여 펩타이드화 시키고 nano flow 의 LCMS 를 이용하여 분석된 데이터를 이용하여 동정된 단백질의 수, 펩타이드 수, PSMs 의 수, C-term K/R ratio 를 분석하여 최적의 추출법을 확인하였다. TNBC 를 needle biopsy 로 획득된 FFPE tissue 를 전단계에서 확인된 최적화된 추출법으로 단백질을 추출하였고 SWATHLC-MS 방법으로 데이터를 획득 후 차별적으로 발현되는 단백질을 확인하였다.

최적의 추출방법을 확인하기 위한 실험에서는 평균 700ug 이상의 단백질을 획득하였다. LC-MS 로 얻은 데이터를 분석하여 각 조건별로 2,000 개 정도의 단백질을 동정하였고 통계분석을 통해 동정률, 재현성, de-crosslink 정도를 확인하여 Test 1 (Extraction buffer EXB, boiling 100 ℃ 에서 20 분, 80 ℃ 에서 4시간, AFA sonication 수행 안함)이 가장 우수하다는 것을 확인할 수 있었다.

단백질 추출의 최적화된 방법으로 core needle biopsy FFPE tissue 147 개에서 단백질을 추출하고 S-trap 으로 얻은 펩타이드를 LC-MS 분석을 진행하여 각 비교군들에 대해 DEP 를 발굴하였다. Non-pCR 과 pCR 을 비교하여 5 개의 단백질을 발굴하였고 non-recurrence 와 Recurrence 를 비교하여 12 개의 양적

i

차이를 갖는 단백질을 발굴하였다. 발굴된 단백질들을 biological function 에 대해 분석하여 항암 치료 반응 및 재발에 대한 기능적인 연관성을 확인하였다.

Keywords: Breast cancer; FFPE, TNBC, LC-MS, SWATH

Abstracti
Contents iii
List of Tables iv
List of Figures ······ v
Abbreviations ······ ix
Introduction 1
Materials and Methods 13
1. Chapter 1 13
i. Sample information 13
ii. Protein Extraction13
iii. Digestion by S-trap15
iv. Nano-LC-ESI-MS/MS Analysis 15
2. Chapter 2 17
i. Sample information 17
ii. Protein Extraction 20
iii. Digestion by S-trap 21
Results 22
Conclusion ······ 35
Discussion ······ 36
References
Abstract

Contents

List of Tables

표 1. FFPE 에서 단백질 추출을	위한 방법 10
표 2. FFPE 조직 샘플에서 최적	의 단백질 추출을 위한 조건
표 3. non-pCR 과 pCR 에 대한	통계 분석으로 획득한 DEPs 목록 27
표 4. non-recurrence 와 recurr	ence 를 통계 분석으로 획득한 DEPs 목록.

List of Figures

그림 1. 단백질체의 특성 1
그림 2. 단백질체학의 중요성 2
그림 3. 의학 및 질량분석 기반의 단백질체의 페러다임 전환
그림 4. 그림 4. 단백질체 분석에 사용 가능한 임상 검체 4
그림 5. Formalin-Fixed, Paraffin-Embedded (FFPE) 시료의 장점 5
그림 6. FFPE tissue 에서 DNA, RNA 추출을 위한 전처리 6
그림 7. FFPE tissue 를 이용한 단백질체 분석의 재현성
그림 8. Formalin-Fixed, Paraffin-Embedded 제작 시 발생하는 protein-protein cross-linking ······ 9
그림 9. HYPER-sol 추출방법으로 FFPE 시료 전처리11
그림 10. Sonicator 의 종류 12
그림 10. 유방암 FFPE tissue 시료 13
그림 11. 최적화된 단백질 추출을 위한 연구 workflow
그림 12. Triple Negative Breast Cancer 의 특징 17
그림 13. Neoadjuvant therapy 의 workflow
그림 14. TNBC FFPE 174 개 시료 정보 18
그림 15. TNBC FFPE 시료 분석을 위한 workflow
그림 16. Extraction 된 단백질들의 Bradford protein assay 결과

그림 17. FFPE 단백질에 대한 LC-MS 결과
그림 18. LC-MS 통계분석 25
그림 19. TNBC FFPE tissue 에 대해 LC-MS 분석으로 획득한 데이터를 통계 분석하여 differential expressed proteins(DEPs)확인
그림 20. 3 종류의 classification 방법에 따른 관해군과 비관해군의 분석 27
그림 21. Random forest 에 의한 관해군과 비관해군의 cluster 분리 28
그림 22. Random forest 에 의한 관해군과 비관해군 classification table 29
그림 23. 3 종류의 classification 방법에 따른 재발군과 비재발군 분석 31
그림 24. Random forest 에 의한 비재발군과 재발군의 cluster 분리 31
그림 25. Random forest 머신 러닝을 이용한 비재발군과 재발군의 Classification
그림 26. 비관해군과 관해군, 재발군과 비재발군에서 공통적으로 발굴된 DEPs
그림 27. SOD2 단백질에 대한 기능 및 The Cancer Genome Atlas (TCGA) 결과
그림 28. 유방암 조직에 대한 SOD2 단백질의 IHC (Immunohistochemistry) 분석
그림 29. PSME1 단백질에 대한 기능 및 The Cancer Genome Atlas (TCGA) 결과
그림 30. 유방암 조직에 대한 PSME1 단백질의 IHC (Immunohistochemistry) 분석····································
그림 31. H1-4 단백질에 대한 기능 및 The Cancer Genome Atlas (TCGA) 결과

분석	nmunohistochemistry)	IHC (In	단백질의	H1-4	대한	조직에	유방암	31.	그림
40		••••		••••		••••••	•••••	•••••	•••••
41		C 분석	multiple I⊦	대한	개에	백질 16	DEP 단	l 32.	그림

Abbreviations

번호	원어	약어
1	Formalin-Fixed Paraffin-Embedded	FFPE
2	Pathologic complete response	pCR
3	Liquid chromatography-mass spectrometry	LC-MS
4	Negative Breast Cancer	TNBC
5	Differential expressed proteins	DEPs
6	Patient derived cell	PDC
7	Patient derived xenograft	PDX
8	Post-transitional modification	PTM
9	Deoxyribonucleic Acid	DNA
10	Ribonucleic Acid	RNA
11	Adaptive-Focused Acoustics	AFA
12	Sodium Dodecyl Sulfate	SDS
13	Dithiothreitol	DTT
14	Iodoacetamide	IAA
15	Triethylammonium bicarbonate buffer	TEAB
16	Electrospray ionization	ESI
17	Normalized collision energy	NCE
18	Gene ontology analysis	GO analysis
19	Estrogen receptor	ER
20	Progesterone receptor	PR
21	Epidermal growth factor receptor2	HER2
22	peptide spectrum matches	PSM
23	Reactive oxygen species	ROS
24	Recurrence	Recur
25	Anthracycline	AC
26	Immunohistochemistry	IHC
27	The Cancer Genome Atlas	TCGA
28	Sequential Window Acquisition of all Theoretical	SWATH

Introduction

단백질체(proteome)은 유전체(genome)보다 약 50 배 더 복잡하다고 알려져 있다. 2000 년대 Human genome project 를 통해서 인간 genome 중 오직 20,000~25,000 개만이 단백질 코딩에 관여한다. [1] 단백질체는 DNA 에서 promoter 와 splicing 등이 이유로 약 10 배 이상의 transcriptome 이 생기게 된다. RNA 에서 protein 이 되는 과정에서 post-transitional modification (PTM)이 생기고 그 조합으로 1,000,000 이상의 proteome 이 존재한다. [2, 3]



Alternative promotes Alternative splicing mRNA editing rost-iransiational mounications

그림 1. 단백질체의 특성, 유전체에서 전사체, 그리고 단백질체가 되는 과정에서 complexity 는 증가하게 된다.



그림 2. 단백질체학의 중요성, morphogenesis 과정에서 유전체는 변화가 없지만 단백질체의 기능적 변화로 인해 phenotype 이 변화됨

다양한 proteome 들은 양적 변화, PTM 의 변화 등을 통해 유기체에서 단백질들간 상호 작용을 하며, 기능적인 변화를 주게 된다.[4]. 그 예로 그림 2 에서 보는 것같이 나비의 morphogenesis 과정에서 애벌레와 성채인 나비의 genome 은 동일하다. 하지만 단백질의 양적, PTM 등의 변화로 기능이 변화하게 되고 애벌레, 번데기, 나비의 phenotype 이 변화하게 된다. 이런 proteome 의 변화들을 대상으로 연구하고 분석하는 방법을 proteomics 라고 하며 medical 적으로 질병의 진단에 필요한 바이오마커 발굴, 신약개발이 가능한 후보물질 발굴 등에 응용된다. [5, 6] 최근에는 정밀의학(precision medicine)분야에서 특정 약물에 예후를 예측하거나, 약물의 표적을 발견할 수 있는 기술로 발전했다.[7-10]



그림 3. 의학 및 질량분석 기반의 단백질체의 페러다임 전환, 질병의 빠른 진단을 위한 바이오마커 발굴 및 확인하는 단백질체 연구가 많이 수행되었지만 최근에는 companion diagnosis 및 drug target discovery 등의 precision Medicine 에 대한 단백질체 연구가 많이 시행 되고 있다.

이렇듯 단백질체 분석은 임상학적으로 응용이 많은 이유는 분석에 사용가능한 임상 검체가 다 양해졌기 때문이다. Plasma, serum 그리고 수술로 획득한 조직 뿐만 아니라 소변, colonoscopy biopsy, FFPE tissue slide, patient derived cell (PDC) / patient derived xenograft (PDX) tissue, Extracellular vesicle, lymphocyte 등이 사용되고 있다.[11-14] 특히 FFPE 는 진단, 수술 시점에서 채취된 검체라는 특징을 가지게 된다.





그림 4. 단백질체 분석에 사용 가능한 임상 검체



그림 5. Formalin-Fixed, Paraffin-Embedded (FFPE) 시료의 장점. A. 병리학적 진단이 완료된 조직을 이용하여 제작된 FFPE tissue block. B. FFPE 시료중 tumor 부위만 확인하여 연구.

FFPE tissue 시료들은 병리과에서 검사가 완료된 상태로 보관되어 있으며, 오랜 시간 동안 실온에서 보관이 가능하다. 또한 시료의 두께를 조절하여 시료의 양을 조절할 수 있다. 마지막으로 frozen tissue 를 확보하지 않고 tumor 조직만 분리하여 분석을 할 수 있다.



그림 6. FFPE tissue 에서 DNA, RNA 추출을 위한 전처리, A. FFPE tissue 에서 회수된 총 DNA 와 RNA. B. 연도별 FFPE tissue 에서 recovery 된 DNA 와 RNA 의 correlation plot. C. FFPE tissue 에서 추출된 DNA 와 RNA 의 수율 간의 상관 관계. D. 보관 중인 다른 FFPE tissue 시료에서 extraction 된 DNA 와 RNA.[15]

이러한 장점을 가진 FFPE tissue 시료들은 기존에는 RNA, DNA 같은 genomics 분야에 많이 사용이 되어왔다.[15-19] 따라서 그림 6 에서와 같이 FFPE tissue 시료에서 genome 을 추출하는 최적화 방법이 존재한다.



그림 7. FFPE tissue를 이용한 단백질체 분석의 재현성. A. Frozen tissue와 FFPE tissue의 재현성 확인을 위한 workflow. B. Frozen tissue와 FFPE에서 정량화 가능한 단백질의 log2로 변환된 단백질양에 대한 scatterplot이다. directly frozen tissue 와 RNAlater 비교 C. immediate formalin-fixed, paraffin-embedded (iFFPE)비교. D. 30분간 보관 후 formalin-fixed, paraffin-embedded (sFFPE)와 비교. E. iFFPE and sFFPE 비교 하였다. [20]

활용도가 다양한 FFPE 시료를 단백질체 연구에서 활용하기 위해 많은 연구가 진행되었다. 우선 Proteomics 연구에 가능한지 알아보기 위해 frozen tissue 와 FFPE tissue 의 재현성 비교하여 FFPE tissue 에서도 충분히 단백질체 연구가 가능하다는 것이 확인되었다.[20] 하지만 FFPE tissue block 을 제작하는 과정에서 protein 과 protein 간에 cross-link 가 이루어지면서 단백질 추출 및 분석에 제한 되었다. 그림 8 을 확인해 보면 생체 분자의 formaldehyde crosslinking 은 2 단계로 발생한다. 첫번째, formaldehyde 는 일반적으로 단백질의 lysin 의 아미노기와 반응하게 된다. 이 반응아미노 기가 methylol 을 형성하여 물분자 하나가 빠져나가면서 Schiff base 를 생성한다. 두번째는 Schiff base 다른 단백질의 아미노기와 결합하여 crosslinking 을 형성하게 된다. [21, 22]



Buffer, pH, Heat, Detergent

그림 8. Formalin-Fixed, Paraffin-Embedded 제작 시 발생하는 protein-protein cross-linking[21, 22].

Protein-protein cross-link 를 제거하기 위해 많은 연구자들이 boiling 의 온도와 시간을 다르게 하거나, pH의 변화, detergent 의 사용, buffer 의 조성들을 다르게 하여 연구를 진행하였다.[22, 23]

표 1. FFPE 에서 단백질 추출을 위한 방법

No.	Ref.	Temp (°C)	Time (min)	рН	SDS (%)	(mM)	Buffer composition		Treatments ^{a)}	Tissue	Maximum yield ^{b)}
	Hara et al. [28]	100	4	6.8	2		125 mM Tris-HCI, 19%	glycerol, 5% 2-ME	:	Brain, gliomas	NA
2	Layfield et al. [29]	90	15	6.8	2.5	20	150 mM Tris-HCl, 8M 10% 2-ME	urea, 20% glycerol,	н	Amyloid deposits	NA
3	lkeda et al. [31]	100+60	20+120	7.6	2		RIPA buffer			Colorectal cancer	12
4 F	Prieto et al. [41]	95	90				Liquid tissue (express	ion pathology)		Colon cancer	NA
5 (Chu et al. [32]	100	30				NDME-PE (Bio-Quick)			Nine different	50
										tissues	
6 (Crockett et al. [43]				0.1		RIPA buffer			Lymphoma cell	NA
										line	
7 1	Palmer-Toy et al. [42]	70	60	8.5	2	20	100 mM AmBic		S	Ear tissues	NA
8	Shi et al. [33]	100 + 60	20+120	7 or 9	2		20 mM Tris-HCl			Renal cancer	1000
1 6	Hwang et al. [44] (1)	94+60	30+180				100 mM AmBic, 30% /	ACN		Prostate cancer	NA
10 H	Hwang et al. [44] (2)	94+60	30+180	7.2	2		RIPA buffer			Prostate cancer	NA
11 F	Rahimi et al. [45]	95	60	9.0			25 mM Tris-HCI			Brain	NA
12 E	Becker et al. [34]	100 + 80	20+120				Oproteome FFPE (Qia	gen)		Seven different	1500
										tissues	
13 .	Jiang et al. [46]	100	30	8.2		65	40 mM Tris-HCl, 6 M (guanidine-HCI	H+S	Liver	NA
14 (Chung et al. [37]	115	15	9.9	1		AgR buffer (Dako), 10	% glycerol	H+P	Prostate	74
15 1	Nirmalan et al. [38]	105	20	6.8	2		100 mM Tris-HCI, 20%	glycerol, 4% 2-ME		Kidney	250
16	Tian et al. [48]	60	120				50% phosphate buffer	, 50% TFE	H+S	Lung	NA
17 5	Sprung et al. [49]	80+60	120 + 60	8.0			50% 100 mM AmBic, 50%	50% TFE	S	Colon adenoma	NA
18 /	Addis et al. [39]	100+80	20+120	8.8	2	200	20 mM Tris HCI			Skel. muscle, liver	700
19 (Ostasiewicz et al. [40]	99	60	8.0	4	100	100 mM Tris-HCI		H+S	Liver	155*
20 /	Azimzadeh et al. [35]	100+80	20+120	8.8	2	200	20 mM Tris-HCl, 1% o	ctylglucoside, 200 mM		Heart	192
21	Nirmalan et al [47]	105+80	20+120			ת	glycine 0 % RaniGest (Water	s) in 50 mM AmBic		Kirlnev	112
a) H, I	homogenization; S, sor	nication; P,	pressure;	۵							
a) n,	indingenization, s, so	lication, r,	pressure,	æ							

b) Maximum yield expressed in mg of proteins per mm³ of tissue; * indicates µg of proteins per mg of tissue; NA, not available.

Proteomics Clin. Appl. 2012, 6, 7–21

가장 최근 개발된 방법을 기준으로 단백질을 solubilization 을 위해 sonication 과정을 진행하고 recovery 하는 연구가 진행되었다. [24] 사용된 sonication 은 probe sonicator 와 AFA sonicator 를 사용되었다. 그림 10 에서 3 가지 종류의 sonicator 를 비교한다. 일반적으로 사용되는 probe sonicator 는 에너지가 시료에 집중되지만 직접적인 접촉으로 시료의 오염이 발생할 수 있다. Water bath sonicator 는 시료에 직접적인 접촉은 없어 오염에 안전하지만 에너지가 분산되어 비효율적이다. AFA(adaptive-focused acoustics) sonicator 는 에너지가 효과적으로 적용되며 시료에 직접적인 접촉이 없어 오염에 안전한 장점이 있다.



그림 9. HYPER-sol 추출방법으로 FFPE 시료 전처리. A. 실험 조건 표. B. 각 조건에서 동정된 peptide. C. 각 조건에서 동정된 단백질. D.각 조건별 동정된 단백질에 대한 Venn diagram, 84%는 동일한 단백질이다.





이번연구에서는 동일한 FFPE 시료를 이용하여 가장 효과적인 protein-protein cross-link 를 제거하여 FFPE tissue 시료로부터 최적화된 단백질 추출 workflow 를 확립한 후에 각 workflow 별 FFPE tissue 시료의 LC-MS 분석으로 특성들을 확인하는 연구를 선행했다. 이를 기반으로 실제 삼중음성 유방암 환자의 치료 전의 FFPE tissue 174 개로부터 최적화된 단백질 추출 workflow 로 단백질 분석을 수행했다. 또한 항암 치료 반응성 및 재발 등 예후를 예측할 수 있는 FFPE 조직 내 단백질 바이오 마커 후보군을 발굴하고 기능에 대한 분석을 진행했다.

Materials and methods

Chapter 1

Sample Information

이번 연구에서 사용된 FFPE tissue시료는 서울아산병원 병리과 이희진 교수님이 제 공하였으며 가로 2 cm, 세로 3 cm, 두께 2 µm의 크기를 가지는 시료였다. 재현성 있는 분석을 위해 동일한 환자에서 생성된 FFPE tissue를 사용하였다.



그림 11. 유방암 FFPE tissue 시료

Protein Extraction

FFPE tissue에서 6가지 조건으로 단백질을 추출하였다. 파라핀을 제거 하기 위해 100% heptane을 500 ul 추가하여 실온 (25°C)에서 1시간동안 배양하였다. Test 1은 Qproteome FFPE kit에서 제시하는 방법을 참조하여 Extraction buffer EXB를 사용 하고 100 °C에서 20분, 80 °C에서 2시간 배양하었다. Test 2은 5% SDS, 50mM buffer를 TEAB(pH 7.55) 추가하고 Adaptive-Focused Acoustics (AFA) Ultrasonication (Covaris, Woburn, MA, USA) 로 peak incident power 175 Watt, burst 200 cycle, duty factor 10%, 온도 4°C로 6분간 solubilization 한후 80 °C에서 1 시간 배양하여 de-crosslink를 과정을 거쳐 단백질을 추출하였다. 동일 조건으로 AFA ultrasonication을 수행하고 추가적인 단백질을 추출하였다. Test 3은 Extraction buffer EXB를 사용하면서 1차 AFA ultrasonication을 수행하고 100 ℃에서 20분, 80 °C에서 2시간 배양한 후 2차 AFA ultrasonication을 진행하였다.



그림 11. 최적화된 단백질 추출을 위한 연구 workflow

Digestion by S-trap

단백질 정량이 완료된 시료에 최종 농도가 20 mM이 되게 1M dithiothreitol(DTT) 을 혼합한 후 Thermomixer (Eppendorf, Hamburg, Germany)를 이용하여 95°C에서 750rpm으로 10분간 disulfide bond를 환원시키는 과정을 수행하였다. 최종 농도가 40 mM이 될 수 있도록 1M iodoacetamide (IAA)를 추가하여 빛이 없는 조건에서 Thermomixer를 이용하여 25°C, 750rpm으로 30분간 alkylation을 수행한 후 12% phosphoric acid가 1.2% phosphoric acid가 될 수 있도록 추가하였다. S-Trap binding buffer (90% methanol, 100 mM Triethylammonium bicarbonate buffer (TEAB) (pH 7.51)) 350 ul을 추가하였다. 모든 시료를 S-Trap™ mini (ProtiFi, Long Island, New York, USA)에 옮긴 후 centrifuge를 4,000g로 30초간 수행하였다. Washing을 위해 S-trap binding buffer 400 ul를 추가하고 centrifuge를 4,000g로 30초간 수행하고 Washing은 3회 수행하였다. S-trap spin column을 새로운 2mL sample tube로 옮긴 후 단백질과 Trypsin/Lys-C mixture (Promega, Madison, WI, USA) 비율이 25:1이 되게 사용하기 위해 trypsin/Lys-C Mix는 50 mM TEAB buffer 에 녹여서 S-trap spin column에 추가하였다. Trypsin/Lys-C가 들어간 S-trap spin column은 흔들리지 않는 상태에서 37 ℃에서 16시간 동안 배양하였다. Peptide elution은 3회에 걸쳐 진행하였으며 첫번째로 50 mM TEAB 80 ul를 추가하고 1,000g로 1분간 centrifuge를 수행하고 두번째로 0.2% formic acid를 80ul 추가한 후 1,000g로 1분간 centrifuge를 수행하였다. 마지막 elution 과정은 0.2% formic acid, 50% acetonitrile 80ul를 추가한 후 4,000g로 1분간 centrifuge를 수행하여 peptide를 용출하였다. Elute는 Cold trap과 결합된 Evaporator (CentriVap Cold Traps, Labconco, Kansas City, MO, USA)을 이용하여 건조시켰다.

Nano-LC-ESI-MS/MS Analysis

Digestion된 peptide는 Dionex UltiMate 3000 RSLCnano system(Thermo Fisher Scientific, Waltham, MA, USA)을 사용하여 분리하였다. Tryptic peptide는 30µL의

15

0.1% formic acid으로 녹여 C18 Pepmap trap column(20mm × 100 µm id, 5 µm, 100 Å, Thermo Fisher Scientific, Waltham, MA, USA)이 장착된 Acclaim™ Pepmap 100 C18 column (500 mm × 75 µm i.d., 3 µm, 100 Å)에 load하였더. 0-40% acetonitrile 0.1% formic acid, 5% DMSO의 농도를 150분 동안 변화를 주었고 50분 동안 column washing 및 안정화 하였다. 분석 시간은 총 200분이었으며 유속은 250nL/min으로 50°C 컬럼 오븐을 이용하였다. Liquid Chromatography는 Electrospray ionization (ESI) source를 사용하여 Q Exactive™ Plus BioPharm에 결합 되어 Mass spectrometry 분석을 수행하였다. Mass spectrometry는 full scan과 20 개의 data-dependent mode로 MS 데이터와 MS/MS 데이터를 획득하였다. MS/MS resolution은 17,500으로 설정하였고 injection time은 50 msec, auto gain control은 165로 설정하였으며 normalized collision energy (NCE)는 27으로 사용하였다.

Chapter 2



Sample Information

그림 12. Triple Negative Breast Cancer의 특징

이번 연구에서 사용된 유방암 환자 FFPE tissue시료는 서울아산병원 병리과 이희진 교수님으로부터 IRB 승인을 받고 제공받았다. Triple Negative Breast Cancer(TNBC) 는 Estrogen receptor (ER), Progesterone receptor (PR), Epidermal growth factor receptor 2 (HER2)의 발현이 없는 유방암이다. 전체 유방암의 15~20%이며, 젊은 여 성에게서 발현되는 빈도가 높다. 또한 basal-like 1, basal-like2, mesenchymal등의 아형을 가지고 있어 진단하기 어렵고 매우 공격적이라 치료도 어렵다는 특징을 가 진다. 현재 TNBC 환자수는 2020년 전세계에 46만명이며, 전이 발생후 1년의 생존 기간을 가진다. 1차 표준치료에 22%가 반응을 보이게 되고 1명당 1768만원 정도의 분담 비용이 발생하는 것으로 알려져 있다. 이번 연구에서 사용된 FFPE 시료는 TNBC로 진단되어 그림 14에서와 같은 Neoadjuvant chemotherapy 후 완전관해군 (pathologic complete response, pCR)로 판명된 122명의 환자로부터 획득된 FFPE 조직 시료와 비관해군(non- pathologic complete response, non-pCR)으로 판명된 52명으로부터 획득한 FFPE 조직시료이다.



*pCR: pathologic complete response

그림 13. Neoadjuvant therapy의 workflow

환자의 연령대는 25세에서 75세까지 정규분포를 이루었고 완전관해군은 30%, 비 관해군은 70% 였다. 또한, 재발된 환자는 37% (64명)이고 재발하지 않은 환자는 63% (110명)에 해당되었다. 총 174개 시료는 Part 1 연구에서 확립된 최적화된 추출 방법을 적용하여 그림 15의 workflow와 같이 진행하였다.



그림 14. TNBC FFPE 174개 시료 정보



그림 15. TNBC FFPE시료 분석을 위한 workflow

Protein Extraction

FFPE tissue에서 단백질을 추출하기 위해 Qproteome FFPE Tissue kit (Qiagen, Germany)를 사용하였으며 Part 1에서 수립된 Method에 따라 진행하였다. FFPE tissue가 붙어 있는 슬라이스에서 면도칼을 이용하여 FFPE tissue를 박리하여 LoBind tube에 옮기고 Heptane 500 ul를 추가하여 실온(25°C)에서 1시간 동안 배양 해 주었다. 100% 메탄올 25ul를 추가하고 10초간 voltexing한 후 9,000g 로 2분간 centrifuge하여 pellet를 확인하였다. Supernatant를 제거한 후 뚜껑을 열고 5분간 수분이 증발되게 하는 과정을 수행하였다. β-mercaptoethanol이 함유된 Extraction buffer EXB Plus 100 ul를 넣고 100 °C에 20분, 80 °C로 4시간 배양하였다. centrifuge를 이용하여 14,000g, 4°C로 15분간 돌려준 후 상층액만 새로운 LoBind tube로 옮겨주었다. 추출된 단백질들에 대해서 Bradford assay (ThermoFisher Scientific, Waltham, MA, USA)를 통해 단백질 정량을 진행하였다.

Digestion by S-trap

단백질 정량이 완료된 시료에 최종 농도가 20 mM이 되게 1M dithiothreitol (DTT) 을 넣어준 후 Thermomixer (Eppendorf, Hamburg, Germany)를 이용하여 95°C에서 750rpm으로 10분간 disulfide bond를 reduction시켜주었다. 최종 농도가 40 mM이 될 수 있도록 1M iodoacetamide (IAA)를 추가하여 빛이 없는 조건에서 Thermomixer를 이용하여 25°C, 750rpm으로 30분간 alkylation을 수행한 후 12% phosphoric acid가 1.2% phosphoric acid가 될 수 있도록 추가하였다. S-Trap binding buffer (90% methanol, 100 mM Triethylammonium bicarbonate buffer (TEAB) (pH 7.51)) 350 ul 추가해 주었다. 모든 시료를 S-Trap[™] mini (ProtiFi, Long Island, New York, USA)에 옮긴 후 centrifuge를 4,000g로 30초간 수행하였다. Washing을 위해 S-trap binding buffer 400 ul를 추가하고 centrifuge를 4,000g로 30초간 수행하였으며 Washing은 3회 수행하였다. S-trap spin column을 새로운 2mL sample tube로 옮겨주었다. 단백질과 Trypsin/Lys-C mixture (Protmga, Madison, WI, USA) 비율은 25:1로 사용하며 trypsin/Lys-C Mix는 50 mM TEAB buffer에 녹여서 S-trap spin column에 추가하였다. Trypsin/Lys-C 이 들어간 S-trap spin column은 흔들리지 않는 상태에서 37 ℃에서 16시간 배양해 주었다. Elution은 3회에 걸쳐 진행하였으며 첫번째로 50 mM TEAB 80 ul를 추가하고 1,000g로 1분간 centrifuge를 수행하였다. 두번째 elution 과정은 0.2% formic acid를 80ul 추가한 후 1,000g로 1분간 centrifuge를 수행하였다. 마지막으로 0.2% formic acid, 50% acetonitrile 80ul를 추가한 후 4,000g로 1분간 centrifuge를수행하였다.. Elute는 Cold trap과 결합된 Evaporator (CentriVap Cold Traps, Labconco, Kansas City, MO, USA)을 이용하여 건조 시켜주었다. 건조가 완료된 시료를 0.1% formic acid로 녹인 후 NanoDrop[™] Onec (ThermoFisher Scientific, Waltham, MA, USA)으로 peptide 정량을 수행하였다..

Nano-LC-ESI-MS/MS Analysis

액체 크로마토 그래피-탠덤 질량 분석법 (LC-MS/MS) 분석 전에 건조된 펩티드 샘플 을 버퍼 A (0.1 % formic acid in HPLC water)으로 녹이고 총 펩티드 농도를 UV/Vis 분광 광도계 (NanoDrop One, Thermo Fisher Scientific)로 흡광도를 측정하였다. 파 장 280 nm, 샘플 유형 옵션이 "1 Abs = 1 mg/mL"로 설정되었다. 그런 다음 iRT-Kit (Biognosys AG, Schlieren, Switzerland)에서 제공하는 iRT 표준을 1/10 부피로 샘플 에 추가하여 공급 업체 지침에 따라 머무름 시간을 보정하였다.[25] 각 샘플의 총량 은 40µg이며 40µL에 용해하였다. 주입된 4 µL의 샘플은 다음 LC-MS/MS 설정과 함 께 SCIEX TripleTOF 5600+ 시스템 질량 분석기를 사용하여 분석하였다. LC 분리를 위해 nanoLC 425 (Eksigent, Dublin, CA, USA)를 Eksigent 마이크로 트랩 C18컬럼 (ChromXP C18CL, 5 µm, 120 Å)과 분석 컬럼으로서 Eksigent 컬럼 (C18-CL, 0.3 x 150 mm, 입자 크기 3 µm, 기공 크기 120 Å)을 함께 컬럼 온도가 40℃로 유지되도 록 하였다. 샘플을 버퍼 A를 사용하여 10µL/min의 유속으로 트랩 컬럼에 로드 하였 다. 10분 후, 펩티드 혼합물은 버퍼 A와 버퍼 B (0.1 % formic acid in HPLC acetonitrile)을 사용하여 57 분에 의해 5µL/분으로 유속으로 농도구배를 통한 분리

21

를 수행하였다. 버퍼 B를 38분 동안 3-25% 변경하고, 5분 동안 25-32% 변경하고, 2 분 동안 32-80% 변경하고, 3분 동안, 80%로 유지하고, 3분 동안 80-3% 변경하고, 8 분 동안 3%로 유지하였다. 개별 샘플의 경우 모든 질량 분석 실행은 SCIEX 기술 노 트에 따라 100개의 가변 창을 사용하여 모든 이론적 질량 스펙트럼 (SWATH) 모드 의 순차 창 수집에서 작동되었다. SWATH 매개 변수는 다음과 같이 설정되었다: Lower m/z limit 400; m/z 상한선 1250; 윈도우 오버랩 (Da) 1.0; CES는 작은 창에 대해 5, 큰 창에 대해 8, 가장 큰 창에 대해 10. MS2 스펙트럼은 고감도 모드에서 2.5ms 동안 100-1500m/z 범위에서 수집되었으며 총 사이클 시간은 2.8초였다. 다 른 MS 파라미터는 다음과 같이 설정되었다: 이온 소스 가스 1 (GS1) 15 psi; 이온 소 스 가스 2 (GS2) 20 psi; 커튼 가스 (CUR) 30 psi; 온도 (TEM) 250 °C; 이온 스프레이 전압 부동 (ISVF) 5500V.

Results

Chapter 1

FFPE 조직에서 단백질 추출

단백질이 포르말린에 고정이 되면 단백질들간 cross-link가 형성된다. 이 crosslink가 조직의 구조를 유지할 수 있게 하지만 아미노기를 가지고 있는 lysin과 arginine에 methylation과 같은 modification을 발생시키며 이러한 modification은 FFPE tissue에서 단백질을 추출하거나 분석에 제한을 주게 된다. 이번 연구에서 protein-protein cross-link를 제거하기 위해 [표 1]과 같이 Extraction buffer, AFA sonication의 사용 유무, 가열 조건들을 비교하였다.

표 2. FFPE 조직 샘플에서 최적의 단백질 추출을 위한 조건

	Deparaffinization	Extraction buffer	AFA sonication	Boiling
Test 1	0	Extraction buffer EXB	Х	100 °C 20min, 80 °C 2hr
Test 2	0	5% SDS, 50mM TEAB (pH 7.55)	0	80 °C 1hr
Test 3	0	Extraction buffer EXB	0	100 °C 20min, 80 °C 2hr

표 2에서처럼 3가지 조건으로 각각 동일 FFPE 조직시료에 대해 단백질 추출을 진 행하였으며 추출된 단백질에 대해 Bradford protein assay로 정량을 진행하였다. Test 1 실험에서는 평균 758 ug의 단백질이 추출되었으며, Test 2에서는 791 ug, Test 3에서는 714 ug이 각각 추출되었다. 3가지 추출방법 모두 700 ug이상의 단백 질들이 추출되어 큰 차이는 보이지 않았다.



Bradford protein assay

그림 16. Extraction 된 단백질들의 Bradford protein assay 결과



그림 17. FFPE 단백질에 대한 LC-MS 결과. A, 각 추출 방법에 대한 venn diagram. B, 방법론 특이적인 단백질들에 대한 GO분석

그림17 A에서처럼 Test 1번, 2번에서 2,200개 이상의 단백질을 동정하였고, Test 3에 서는 단백질 1,800개로 동정된 단백질 수가 감소하였다. 동정된 단백질들을 Venn diagram으로 비교하였을 때 72%에 해당되는 단백질들이 3가지 조건에서 공통적으 로 동정되었다. 그리고 약 5%의 단백질들이 3가지 방법별 특이적으로 동정된 것을 확인하였다. 이들 방법별 특이적인 단백질에 대해 Gene ontology 분석을 수행한 결 과 단백질들의 cellular location에 의한 차이는 없음을 확인하였다.



그림 18. LC-MS 통계분석, A. The number of identification Peptide, B. The number of identification PSMs, C. Peptide C-Term K/R ratio

각 방법에 따라 동정된 단백질들에 대해 Box plot 형식으로 동정된 peptide 수, 동 정된 peptide spectrum matches (PSMs)수, 그리고 de-crosslink의 정도를 반영하는 peptide C-term lysin/arginine ratio를 분석하였다. 그림18 에서처럼 동정된 peptide 수와 PSMs의 수에 대해 통계적으로 차이는 없는 것을 확인하였다. 그러나 재현성 부분에서 Test 1과 Test 3이 Test2보다 보다 재현성이 우수하다는 것을 확인하였다. 때 아미노기에 Protein-protein cross-link가 형성되었을 lysine의 methylation(+14Da), methylene(+12Da), methylol(+30Da)의 modification이 발생 하게 된다. Trypsin/Lys-C을 이용하여 digestion 후 C-terminal의 lysine과 arginine의 비를 확인하였을 때 de-crosslink이 균일하게 lysine에서 일어났다면 이 비율이 1에 가까운 수가 되고 de-crosslink이 bias를 가지고 발생했다면 이 비율이 1보다 적게 된다. 3가지 방법에 대해 C-terminal의 K/R의 ratio를 확인하였을 때 Test1과 Test 3 은 median값이 약 1이며 Test 2는 0.9이므로 Test 1과 Test 3이 Test2에 비해 decrosslink 반응이 우수하다고 볼 수 있다. 종합적으로 동정률, 재현성, De-cross linking정도가 Test 1이 가장 우수하다는 결론을 도출하였다.

Chapter 2

Part 1 연구에서 최적화된 단백질 추출 workflow를 적용하여 TNBC 환자의 FFPE 시 료에 대해 단백질체 분석을 수행하였다. 또한 LC-MS 결과에 대한 정량적 통계 분석 을 수행하여 항암 치료 반응성 및 재발 등 예후를 예측할 수 있는 바이오 마커 후보 군을 발굴하고 해당 단백질들에 대한 기능에 대하여 분석하였다. LC-MS 분석 결과 총 5,547개의 단백질 그룹이 정량에 사용되었고 이들 단백질을 이용하여 통계 분석 을 진행하여 확인한 differential expressed proteins (DEPs)에 대한 volcano plot의 결과는 그림19으로 나타내었다.



그림 19. TNBC FFPE tissue에 대해 LC-MS 분석으로 획득한 데이터를 통계분석하 여 differential expressed proteins(DEPs)확인, A. non-pCR과 pCR의 발현량을 비 교하였으며 pCR에서 증가되는 단백질 4개와 non-pCR에서 증가되는 단백질 1개를 확인, B. non-recurrence와 recurrence를 비교하여 recurrence에서 증가 되는 단백 질 4개, non-recurrence에서 증가하는 단백질 8개를 확인 표 3. non-pCR과 pCR에 대한 통계분석으로 획득한 DEPs 목록. 붉은 글씨는 pCR에 서 증가한 단백질, 푸른 글씨는 non-pCR에서 증가한 단백질

`Gene	Description	pCR/Non-pCR	p-value
CORO1A	Coronin-1A	1.507712	0.000463
`SOD2	Superoxide dismutase [Mn], mitochondrial	1.389375	0.001898
`SRSF1	Serine/arginine-rich splicing factor 1	1.220117	0.005837
`PFN1	Profilin-1	1.237027	0.00118
`KRT14	Keratin, type I cytoskeletal 14	0.616044	0.007276

치료 반응에 대한 5개의 단백질들을 이용하여 Neoadjuvant therapy의 반응성을 구 분하기 위해 Principal component analysis (PCA), Partial Least square discriminant analysis (PLS-DA) 및 Self-organizing map (SOM) 분석을 수행하여 174명의 시료가 관해군과 비관해군이 잘 구별되는 것을 확인하였다.



그림 20. 3종류의 classification 방법에 따른 관해군과 비관해군의 분석

그림 21에서는 174개 시료를 관해군과 비관해군에서 차이나는 단백질 5개들을 이 용하여 hierarchical clustering analysis를 수행하였다. 그 결과 관해군과 비관해군에 서 cluster가 분리되는 것을 확인할 수 있었다.



그림 21. Random forest에 의한 관해군과 비관해군의 cluster 분리

174개 시료에 대해 5개의 단백질을 이용하여 classification을 수행한 결과 그림 22 에 확인되듯이 비관해군 122명 중 109명을 구분 할 수 있었다. 이는 89.3%의 정확 도를 확인할 수 있었다. 이 와 같은 결과는 Neoadjuvant therapy 전에 비관해군을 89.3%의 정확도로 예측할 수 있다는 점에서 큰 의미를 가진다.



그림 22. Random forest에 의한 관해군과 비관해군 classification table

표 4. non-recurrence와 recurrence를 통계 분석하여 획득한 DEPs 목록. 붉은색 글씨는 recurrence에서 증가되는 단백질 4개, 푸른 글씨는 non-recurrence에서 증 가하는 단백질

`Gene	Description	Recur. O/Recur X	p-value
SERPINA1	Alpha-1-antitrypsin	1.463614	0.003214
POSTN	Periostin	1.438456	0.004152
TUBB4A	Tubulin beta-4A chain	1.391903	0.000889
COL6A3	Collagen alpha-3(VI) chain	1.296746	0.008966
GNA13	Guanine nucleotide-binding protein subunit alpha-13	0.845361	0.009296
`H3C15	Histone H3.2	0.835827	0.002377
H2AC21	Histone H2A type 2-B	0.806119	0.005785
NME2P1	Putative nucleoside diphosphate kinase	0.764888	0.005351
FERMT3	Fermitin family homolog 3	0.741703	0.001526
`PSME1	Proteasome activator complex subunit 1	0.741228	0.001823
SOD2	Superoxide dismutase [Mn], mitochondrial	0.725101	0.003501
`HLA-A	HLA class I histocompatibility antigen, A alpha chain	0.714604	0.006457

치료 전의 조직 시료에서 치료 예후인 재발군과 비재발군에서 확인된 DEPs 12개의 단백질들을 확인하였다. 12개의 단백질들을 이용하여 재발군과 비재발군을 구분하 기 위해 Principal component analysis (PCA), Partial Least square discriminant analysis (PLS-DA) 및 Self-organizing map (SOM) 분석을 수행하였다. 그 결과 대체 로 두 그룹은 잘 구분되지 않는 것을 확인하였다.



그림 23. 3종류의 classification 방법에 따른 재발군과 비재발군 분석

174개의 시료를 재발군과 비재발군의 DEPs 12개의 단백질들을 이용하여 hierarchical clustering analysis를 수행하였다. 그 결과 재발군과 비재발군에서 cluster가 분리되는 것을 관찰할 수 있었다. 재발군 보다 비재발군에서 cluster가 더 잘 형성되는 것을 확인되었다.



그림 24. Random forest에 의한 비재발군과 재발군의 cluster 분리

Random forest 머신 러닝 방법을 이용해 classification 분석을 진행하였다. 12개의 DEPs 단백질을 이용한 모델을 통하여 총 174개 시료에서 128개의 비재발군 중에 119개의 비재발군을 92.97%의 정확도로 구별하는 것으로 확인되었다. 이 모델은 TNBC 환자의 항암 치료 전에 항암 치료의 예후가 긍정적일 가능성에 대한 예측이 가능하기 때문에 임상적으로 매우 중요하다.



그림 25. Random forest 머신 러닝을 이용한 비재발군과 재발군의 Classification

그림 26에서 관해군과 비관해군의 비교, 재발군과 비재발군의 비교에서 공통적으 로 확인되는 DEP 3개에 대해 box plot분석을 수행하였다. H1-4, PSME1, SOD2는 비관해군에서 감소하고 관해군에서는 증가되는 것을 확인하였다. 또한 3개의 단백 질은 재발군에서는 감소하고 비재발군에서는 증가하는 것을 확인하였다.



그림 26. 비관해군과 관해군, 재발군과 비재발군에서 공통적으로 발굴된 DEPs

Conclusion

본 연구자는 Part 1 연구를 통해 2 µm 두께의 FFPE tissue slice 에서 3 종류의 다른 조건에서 700 µg 이상의 단백질을 추출하였고 LC-MS 분석으로 2,000 개이상의 protein group 을 동정할 수 있었다. 해당 결과를 이용할 경우 기능적인 단백질체 분석이 가능한 결과로 활용할 수 있을 것으로 사료된다. 또한 LC-MS 분석을 통해 얻은 데이터를 통계분석을 통해 시료 전처리의 재현성 및 동정율, De-cross linking 의 정도가 Test 1 조건 (Extraction buffer EXB, boiling 100 ℃에서 20 분, 80 ℃에서 4 시간, AFA sonication 수행 안함)이 가장 우수한 것으로 확인되었다. Test 1 조건에서 AFA sonication 만 추가하여 진행한 Test 3 조건 (Extraction buffer EXB, boiling 100 ℃에서 20 분, 80 ℃에서 4 시간, AFA sonication 수행)은 재현성이나 de-crosslinking 정도는 Test 1 과 의미 있는 차이를 가지지 않았으나 동정률에서 Test3 가 Test1 보다 약 20% 낮았다. 이는 AFA sonication 의 강도가 강하여 de-crosslinking 뿐만 아니라 단백질들까지 분해된 것으로 추측된다. 이를 해결하기 위해 추가적으로 AFA sonication 의 강도를 낮추는 method 조건을 확인하는 실험을 추가로 진행하면 추가적인 결론을 내릴 수 있을 것으로 기대한다.

Part 2 에서의 연구에서는 Part 1 에서 확립된 FFPE 시료 전처리 기법을 적용하여 174 명의 TNBC 환자로부터 획득한 FFPE tissue 시료에 대해 단백체 분석을 수행하였다. 이를 활용하여 항암 치료 예후 바이오마커 후보군 발굴연구를 수행하였으며 관해군과 비관해군 비교에서 5 개의 단백질을 확인 하였고 재발군과 비재발군 비교에서 12 개의 단백질을 확인 할 수 있었다. 총 16 개 (SOD2 중복)는 항암 치료 전후의 바이오마커, 항암 치료 이후의 재발들의 예후 예측할 수 있는 바이오마커로 가능성이 있을 것으로 예상된다. 또한 관해군과 비관해군 비교, 재발군과 비재발군에서 공통적으로 3 개의 단백질을 확인하였다. H1-4, PSME1, SOD2 는 비관해군에서 감소하고 관해군에서는 증가되는 것으로 확인되었으며, 재발군과 비재발군의 비교에서는 반대로 비재발군에서 증가하고 재발군에서 감소하는 것으로 확인되었다. 3 가지 단백질 모두 TCGA 분석으로 발현이 적으면

35

유방암에 대해 나쁜 예후를 가지는 것으로 확인이 되었다. 이 3 개 단백질들은 TNBC 치료 예후 및 치료 반응성 예측을 위한 플렛폼 개발에 확용이 가능할 것으로 보인다.

Discussion

본 연구에서 FFPE 조직을 제조하는 과정에 발생하는 protein-protein cross-link를 효과적으로 제거하고 단백질 추출을 최적화하기 위해 Buffer의 종류, De-cross linking time, AFA sonication의 적용 여부를 다르게 하여 변화시켰다. 총 3가지 조건 의 method를 operation하였고 각 실험에 따라 추출된 단백질을 trypsin으로 peptide화 하여 LC-MS분석을 통해 데이터를 획득하였다. 획득한 데이터를 이용하 여 통계분석을 통해 동정된 단백질의 수, 동정된 peptide의 수, Mapping된 peptide spectrum의 수, C-Term K/R의 ratio를 비교하였다. 최종적으로 Extraction buffer EXB를 사용하며, 100 °C에서 20분, 80 °C에서 4시간 가열한 Test 1 조건이 재현성과 동정률, de-cross liking 정도가 우수하다고 판단하여 최종 선정하였다.

FFPE tissue 시료에서 최적의 단백질 추출 workflow를 적용하여 TNBC 진단을 받고 Neoadjuvant 치료를 받은 174명의 환자들로부터 획득한 FFPE시료에 대해 단백체 분석을 수행하였다. 174개의 시료는 완전 관해군과 비관해군으로 나뉘며 재발 및 비재발군으로 나뉘었으며 LC-MS 분석을 진행하여 각 비교군들에 대해 DEP들을 선 정하였다. Non-pCR과 pCR을 구분할 수 있는 5개의 단백질들을 발굴하였고 nonrecurrence와 Recurrence를 비교하여 12개의 양적 차이를 갖는 단백질들을 발굴하 였다. 발굴된 단백질들을 biological function에 대해 분석하여 항암 치료 반응 및 재 발에 대한 기능적인 연관성을 확인하였다.

특히 H1-4, PSME1, SOD2 단백질들은 비관해군에서 감소하고 관해군에서 증가되는 것으로 확인되었다. 재발군과 비재발군의 비교에서는 반대로 재발군에서 감소하고 비재발군에서 증가하는 것으로 확인되었다. 3개의 단백질 들에 대해 IHC와 MS 기 반 다중형광염색 기반의 TNBC 치료 예후 예측 및 치료 반응성 예측을 위한 플랫폼 개발에 활용 가능할 것으로 보인다.

36

Superoxide dismutase [Mn], mitochondrial / SOD2



그림 27. SOD2 단백질에 대한 기능 및 The Cancer Genome Atlas (TCGA) 결과 A. SOD2의 Gene Ontology 분석. B. SOD2의 TCGA 분석 파란선이 low expression으로 핑크색의 high expression보다 예후가 안 좋은 것으로 확인

SOD2 (Superoxide dismutase [Mn], mitochondrial) 단백질은 일반적으로 세포내에 서 생성되고 독성이 있는 superoxide anion radical을 파괴하는 역할을 한다. 그리고 drug에 대한 response와 hydrogen peroxide response에 대한 기능을 가지고 있으 며 SOD2가 낮게 발현이 되면 유방암을 비롯하여 다양한 암에 대해 poor prognosis 를 나타내는 것으로 알려져 있다. [26-29] SOD2는 두 종류의 항체 (HPA001814와 CAB002013)를 이용하여 Tissue microarray 분석 결과를 분석한 결과 유방암 조직에 서 검출이 되는 것으로 나타났다.



그림 28. 유방암 조직에 대한 SOD2 단백질의 IHC (Immunohistochemistry) 분석.

Proteasome activator complex subunit 1 / PSME1



그림 29. PSME1 단백질에 대한 기능 및 The Cancer Genome Atlas (TCGA) 결과 A. PSME1의 Gene Ontology 분석. B. PSME1의 TCGA 분석 파란선이 low expression으로 핑크색의 high expression보다 예후가 안 좋은 것으로 확인

PSME1 (Proteasome activator complex subunit 1) 단백질은 일반적으로 Immunoproteasome 생성과 관련이 있으며 효율적인 항원 처리에 기능을 수행하는 것으로 알려져 있다. PA28 활성화제 복합체는 proteasome의 절단 패턴을 변경하여 generation of class I binding peptides 생성을 향상시킨다. 또한, PSME1 단백질은 G1/S transition 과정에 관여하고 또한 stem cell differentiation에 관여된 기능을 가 지고 있다.[30] PSME1의 낮은 발현량은 유방암에 있어서 poor한 prognosis를 보이 는 것으로 나타났으며 이 결과는 TNBC 항암 치료 반응과 예후와 관련된 결과와 동 일한 것으로 확인된다. [31, 32]



그림 30. 유방암 조직에 대한 PSME1 단백질의 IHC (Immunohistochemistry) 분석.



그림 31. H1-4 단백질에 대한 기능 및 The Cancer Genome Atlas (TCGA) 결과 A. H1-4의 Gene Ontology 분석. B. H1-4의 TCGA 분석 파란선이 low expression 으로 핑크색의 high expression보다 예후가 안 좋은 것으로 확인

H1-4 (Histone H1) 단백질은 chromatin fiber로 DNA와 nucleosome을 연결해주는 거대 분자구조이다. Histone H1은 nucleosome chains의 응축하는데 필요하다. 또한 chromatin remodeling, nucleosome spacing and DNA methylation을 통해 개별 gene transcription의 조절자 역할도 하는 것으로 알려졌다. H1-4 단백질도 위의 2 개의 단백질과 같이 낮은 발현량은 유방암에 있어서 poor한 prognosis를 보이는 것 으로 나타났다.



그림 31. 유방암 조직에 대한 H1-4 단백질의 IHC (Immunohistochemistry) 분석.

'Gene	Description	pCR/Non-pCR	p-value
CORO1A	Coronin-1A	1.507712	0.000463
SOD2	Superoxide dismutase [Mn], mitochondrial	1.389375	0.001898
SRSF1	Serine/arginine-rich splicing factor 1	1.220117	0.005837
PFN1	Profilin-1	1.237027	0.00118
'KRT14	Keratin, type I cytoskeletal 14	0.616044	0.007276
Gene	Description	Recur. O/Recur X	p-value
SERPINA1	Alpha-1-antitrypsin	1.463614	0.003214
POSTN	Periostin	1,438456	0.004152
TUBB4A	Tubulin beta-4A chain	1.391903	0.000889
COL6A3	Collagen alpha-3(VI) chain	1.296746	0.008966
GNA13	Guanine nucleotide-binding protein subunit alpha-13	0.845361	0.009296
'H3C15	Histone H3.2	0.835827	0.002377
H2AC21	Histone H2A type 2-8	0.806119	0.005785
'NME2P1	Putative nucleoside diphosphate kinase	0.764888	0.005351
FERMT3	Fermitin family homolog 3	0.741703	0.001526
'PSME1	Proteasome activator complex subunit 1	0.741228	0.001823
SOD2	Superoxide dismutase [Mn], mitochondrial	0.725101	0.003501
HLA-A	HLA class I histocompatibility antigen, A alpha chain	0.714604	0.006457

그림 32. DEP 단백질 16개에 대한 multiple IHC 분석.

항암 치료에 대한 관해군과 비관해군을 비교하여 확인된 5개의 DEPs와 치료 예후에 대한 재발행군과 비재발행군을 비교하여 확인된 12개 DEPs 중 중복되는 SOD2를 감안하여 총 16개의 단백질들에 대해 그림 32와 같이 multiple IHC 분석법을 수행 하여 진단 가능성에 대한 추가 연구를 수행할 수 있을 것으로 기대한다.

References

- 1. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
- Aebersold, R., et al., *How many human proteoforms are there?* Nat Chem Biol, 2018. 14(3): p. 206-214.
- 3. Wold, F., *In vivo chemical modification of proteins (post-translational modification).* Annual review of biochemistry, 1981. **50**(1): p. 783-814.
- Pandey, A. and M. Mann, *Proteomics to study genes and genomes.* Nature, 2000. 405(6788): p. 837-846.
- Srinivas, P.R., et al., *Proteomics for cancer biomarker discovery.* Clinical chemistry, 2002. 48(8): p. 1160-1169.
- Burbaum, J. and G.M. Tobal, *Proteomics in drug discovery*. Current Opinion in Chemical Biology, 2002. 6(4): p. 427-433.
- Ahn, H.S., et al., Differential Urinary Proteome Analysis for Predicting Prognosis in Type 2 Diabetes Patients with and without Renal Dysfunction. Int J Mol Sci, 2020. 21(12).
- Chambliss, A.B. and D.W. Chan, *Precision medicine: from pharmacogenomics to pharmacoproteomics.* Clinical proteomics, 2016.
 13(1): p. 1-9.
- 9. Prasad, B., et al., *The promises of quantitative proteomics in precision medicine.* Journal of pharmaceutical sciences, 2017. **106**(3): p. 738-744.
- Uzozie, A.C. and R. Aebersold, *Advancing translational research and precision medicine with targeted proteomics.* Journal of proteomics, 2018.
 189: p. 1-10.
- 11. Issaq, H.J., Z. Xiao, and T.D. Veenstra, *Serum and plasma proteomics.* Chemical reviews, 2007. **107**(8): p. 3601-3620.
- 12. Chaurand, P., et al., *Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections.* The American journal of pathology, 2004. **165**(4): p. 1057-1068.

- 13. Decramer, S., et al., *Urine in clinical proteomics.* Molecular & cellular proteomics, 2008. **7**(10): p. 1850-1862.
- 14. Chia, S., et al., *Phenotype-driven precision oncology as a guide for clinical decisions one patient at a time.* Nature communications, 2017. 8(1): p. 1-12.
- 15. Patel, P.G., et al., *Preparation of formalin-fixed paraffin-embedded tissue cores for both RNA and DNA extraction.* Journal of visualized experiments: JoVE, 2016(114).
- 16. Schweiger, M.R., et al., *Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number-and mutation-analysis.* PloS one, 2009. **4**(5): p. e5548.
- Mathieson, W. and G. Thomas, Using FFPE tissue in genomic analyses: advantages, disadvantages and the role of biospecimen science. Current Pathobiology Reports, 2019. 7(3): p. 35-40.
- Bass, B.P., et al., A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen? Archives of pathology and laboratory medicine, 2014. 138(11): p. 1520-1530.
- Kerick, M., et al., *Targeted high throughput sequencing in clinical cancer* settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. BMC medical genomics, 2011.
 4(1): p. 1-13.
- Bennike, T.B., et al., *D Comparing the proteome of snap frozen, RNAlater preserved, and formalin-fixed paraffin-embedded human tissue samples.* EuPA Open Proteomics, 2016. **10**: p. 9-18.
- Lee, D.K., et al., Removing Formaldehyde-Induced Peptidyl Crosslinks Enables Mass Spectrometry Imaging of Peptide Hormone Distributions from Formalin-Fixed Paraffin-Embedded Tissues. Angewandte Chemie International Edition, 2020. 59(50): p. 22584-22590.
- 22. Hoffman, E.A., et al., Formaldehyde crosslinking: a tool for the study of

chromatin complexes. Journal of Biological Chemistry, 2015. **290**(44): p. 26404-26411.

- Azimzadeh, O., et al., Formalin-Fixed Paraffin-Embedded (FFPE) Proteome Analysis Using Gel-Free and Gel-Based Proteomics. Journal of Proteome Research, 2010. 9(9): p. 4710-4720.
- 24. Marchione, D.M., et al., *HYPERsol: flash-frozen results from archival FFPE tissue for clinical proteomics.* bioRxiv, 2019: p. 632315.
- 25. Escher, C., et al., *Using i RT, a normalized retention time for more targeted measurement of peptides.* Proteomics, 2012. **12**(8): p. 1111-1121.
- Chang, B., et al., SOD2 deregulation enhances migration, invasion and has poor prognosis in salivary adenoid cystic carcinoma. Scientific reports, 2016. 6(1): p. 1-10.
- Xu, Z., et al., SOD2 rs4880 CT/CC genotype predicts poor survival for Chinese gastric cancer patients received platinum and fluorouracil based adjuvant chemotherapy. American journal of translational research, 2015.
 7(2): p. 401.
- Miar, A., et al., Manganese superoxide dismutase (SOD2/MnSOD)/catalase and SOD2/GPx1 ratios as biomarkers for tumor progression and metastasis in prostate, colon, and lung cancer. Free Radical Biology and Medicine, 2015. 85: p. 45-55.
- 29. Termini, L., et al., *SOD2 immunoexpression predicts lymph node metastasis in penile cancer.* BMC clinical pathology, 2015. **15**(1): p. 1-8.
- Yamada, T., et al., Expression of stem cell markers as useful complementary factors in the early detection of urinary bladder carcinogens by immunohistochemistry for γ-H2AX. Archives of Toxicology, 2021. 95(2): p. 715-726.
- 31. Lino, M.A.M., et al., *Comparative proteomic profiling of triple-negative breast cancer reveals that up-regulation of RhoGDI-2 is associated to the inhibition of caspase 3 and caspase 9.* Journal of proteomics, 2014. **111**: p. 198-211.

32. Lou, S., et al., *High nuclear expression of proteasome activator complex subunit 1 predicts poor survival in soft tissue leiomyosarcomas.* Clinical sarcoma research, 2016. **6**(1): p. 1-7.

Abstract

Optimization of sample preparation for formalin-fixed paraffinembedded tissue proteome

Formalin-Fixed Paraffin-Embedded (FFPE) tissue has the advantage that it has completed pathological examination and is being stored in the pathology department, and the sample volume can be adjusted through thickness adjustment, and only tumor tissue can be selected.

Although the FFPE tissue sample has many advantages, it forms a protein-protein crosslink during the production process and is limited to protein extraction and proteomics analysis. In this study, we confirmed the optimized protein extraction method using the same FFPE tissue sample, extracted the protein from 147 TNBC FFPE samples by the optimized method, performed LC-MS analysis, and treated. We excavated biomarkers for prognosis and recurrence.

Proteins were extracted from 2 um FFPE tissue slices under a total of three conditions, such as extraction buffer, presence / absence of AFA sonicator, and boiling, then protein was quantified. The extracted protein was peptided using the S-Trap digestion method, and the number of identified proteins, number of peptides, number of PSMs, and C-term K / R ratio were analyzed using the data analyzed using Nanoflow LCMS. Then, the optimum method for extracting the protein from the FFPE sample was confirmed. Proteins were extracted from the FFPE tissue obtained by needle biopsy for TNBC by the optimized extraction method confirmed

in the previous step, and after data were obtained by the SWATH method, proteins that were discriminatively expressed were confirmed.

In an experiment to confirm the optimum extraction method, an average of 700 μ g or more of protein was obtained. Analyzing the data obtained by LCMS, identifying about 2,000 proteins under each condition, confirming the identification rate, reproducibility, and de-crosslink level through statistical analysis, Test 1 (Extraction buffer EXB, boiled at 100 °C for 20 minutes and 80 °C for 4 hours, without AFA sonication) was confirmed to be the best.

Proteins were extracted from 147 core needle biopsy FFPE tissues by an optimized method of protein extraction, and the peptides obtained by S-trap were subjected to LC-MS analysis, and DEPs were excavated for each comparative group. Five proteins were excavated by comparing Non-pCR and pCR, and 12 proteins with a quantitative difference were excavated by comparing non-recurrence and Recurrence. The excavated protein was analyzed for biological function and its functional association with anticancer treatment response and recurrence was confirmed.

Keywords: Breast cancer; FFPE, TNBC, LC-MS