

Human metabolome 분석 기술별 dataset 통합 연구법

Integration of datasets from different analytical techniques to assess the impact of nutrition on human metabolome

이영경 | 공정기술연구단

Rhee, Young Kyoung | Processing Technology Research Group

서론

사람의 장내에는 최소 100여종의 미생물이 수백조개가 살고 있다. 미생물 분포가 가장 많은 곳 중의 하나인 colon에서 bacteria의 농도는 $10^{11} \sim 10^{12}$ /ml로 측정되고 있다. 이 미생물 ecosystem은 미생물이 거주하는 인체의 면역 능력, 영양소 대사, 혈관 형성 자극, 장관면역 조절, 체내 지방축적 조절 등의 여러 가지 중요한 기능에 영향을 미친다. 성인 소화기관의 미생물 층은 전통적인 미생물 배양법과 최근의 small unit(SSU) rDNA sequencing법을 이용해 많이 연구되어 있다.

소화기와 연관된 질병의 예방이나 치료는 복잡한 소화 과정을 포함하여 전체 기전을 이해하는 것이 중요하다. 그러나 식이, 미생물층, 인체간의 상호작용은 알려진 것이 거의 없다.

본문은 혈액, 소변, 대변 등에서 소화대사산물, 식이-장내세균총 또는 인체-장내세균총 상호작용

의 산물에 존재하는 성분을 탐지하거나 분리하는 연구법들에 대하여 설명하고자 한다.

식이가 대사체에 미치는 영향

식품 내의 수천개 화합물은 소화·대사 같은 생리적 작용과 미생물총의 개입으로 혈액, 소변, 대변, 타액 등의 체액에 무수히 많은 대사체를 남긴다. 일상적인 식이조건에서는 15,000개 이상의 영양소와 8,000개 이상의 비영양적 요소(식이섬유, 향산화제, prebiotics, probiotics 등)에 접할 수 있다. 일례로 우유는 200여개의 서로 다른 oligo-saccharides와 10,000여개의 탐지 가능한 식용 식물 대사체, 800여개의 비영양성 phytochemicals를 함유하고 있다. 이 물질들 중 사용되지 않거나 남는 양이 체액(특히 소변)에서 발견된다. 또 와인, 홍차, 커피, 과일 등은 소변과 혈액의 바이

오마커 검색을 통해 섭취여부를 확인할 수 있다. 8-isoprostaglandin F2alpha 같은 지질과산화물 또는 malonaldehyde나 glutathione은 산화적 스트레스 마커의 산물들이다.

일부 식이 성분은 장내 균총을 특징짓는 미생물에 매우 직접적인 영향력을 행사하여 양적·질적 조절을 유도한다. 미생물의 대사산물은 결과적으로 중심에 적합한 환경이 된 인체내에서 생리적으로 활성이 있는 화합물로 작용을 한다. 여기에 인체의 상피세포는 미생물 발효산물을 에너지원으로 사용할 수 있도록 적응한다. 단쇄 지방산(short chain fatty acid, SCFA)은 장관에서 이뤄지는 생화학적-생리학 과정에 분명한 영향을 미칠 수 있고, 일부 중요한 생물학적 기능이 이뤄지도록 한다. 그러나 대사산물 생성과 장관 건강에서 발견되는 개인차에 대해 미생물 균총 조성이 갖는 영향은 알려진게 거의 없고, 다만 특정 대사산물과 대사질환과의 관계만이 알려져 있다.

대사 프로파일링은 장내 미생물 균총 요소간의 복잡한 상호 작용을 이해하는데 큰 장점이 있고, 특정 영양요소가 장내 균총 조성 변화에 미치는 영향을 설명한다. 특히, 식이 형태와 관련한 대사적 특징을 확인하는 연구는 도전해 볼 만한 가치가 있다. 사실 체액에서 관찰할 수 있는 대사체 전체 또는 metabotype이라고 불리는 개인의 대사체 phenotype 전체는 환경, 약물, 식이, 생활습관, 유전 등 다양한 외인성/내인성 인자에 영향을 받는다. 특히 식이 성분 중에서는 영양성 성분 보다는 비영양성 성분이 더 큰 영향을 미친다고 여겨진다. 비영양성 성분은 섭취한 식품을 추정할 수 있는 바이오마커가 되는 대사체 프로파일에 중요한 영향을 준다.

인체와 장내 미생물의 대사적 상호작용과 cell-cell communication

포유동물과 미생물의 상호작용에 따라 Insulin 저항성, Crohn병, 과민성대장증상, 식품 알리지, 위염, 위궤양, 비만, 심혈관 질환, 소화기 암 등 여러 질병의 원인이 될 수 있다. 장내 미생물의 활동은 매우 특이적이어서 Bifidobacterium의 경우, 면역계 발달과 장 기능 유지에 중요한 것으로 알려져 있다. 장내 미생물은 인체와 밀접한 작용을 통해 metabotype을 결정하기 때문에 개인 맞춤형 건강관리 솔루션의 해법이 될 수 있다. 또한 식이는 장내 미생물 형성에 중요한 영향을 미쳐 metabolic signaling network 구축에 결정적인 역할을 한다.

인체와 장내 미생물이 공조하여 식품과 제노바이오틱스(이물질 또는 장내 미생물이 생산한 화합물)을 전환시켜, 다수의 작은 분자들을 생산하고, 이 분자들은 인체 세포와 공생하는 장내 미생물 사이에 정보를 전달하는 중요한 역할을 수행한다. 여기에는 저분자 대사물질, peptides, 단백질을 이용한 신호전달이 포함되고, 면역매개경로를 통해 간접적으로 기능 발현이 진행될 수 있다. 장내 미생물의 대사 물질은 인체 건강 상태에 영향을 미칠 수 있고, 체액에서 그것이 탐지된다는 것은 질병 발생 지표나 장내 미생물총에 이상이 생긴 것으로 여겨질 수 있다(Fig 1).

인체 GI tract에는 부위마다 다양한 조성으로 장내 미생물이 존재한다. GI tract의 각 부위에서 장내 미생물과 인체가 그렇듯이 서로 다른 종류의 미생물 간에도 화학물질에 의한 상호 작용을 한다.

대사물질은 모든 생체에 존재하는 복잡한 생합성 또는 이화학적 경로의 산물이거나 부산물이다.

그러므로 건강한 생체 시스템은 중간 대사체의 농도를 거의 일정하게 유지하려고 한다.

분변 추출물은 포유류와 장내 생태환경 간의 복잡한 대사 작용을 이해하기 위한 연구에서 사용하는 가장 흔한 시료이고, 여기서 단쇄 지방산, 아미노산, 유기산, nucleosides, nucleotides, polyamines, 페놀 화합물, 담즙산, 지질 분획 관련 대사 프로파일과 같이 소화기 질환 발생 여부를 확인할 수 있는 정보를 찾을 수 있다.

Grider와 Piland(2007)에 의하면 단쇄 지방산은 장관 수송을 자극하고, *in vitro* 장상피 시스템 연구에서 serotonin 분비를 8~10배 가량 높인다는 것을 확인하였다. 또, Musso 등(2011)은 장내 미생물이 생산하는 가장 중요한 대사물질이 단쇄 지방산이고, 공생하는 인체의 열량 이용성, 인체-미생물간의 신호전달, 장내 미생물 조성 변화에 의한 장내 pH 조절, 장 운동성, 상피세포 증식 등에 영향을 미친다고 하였다.

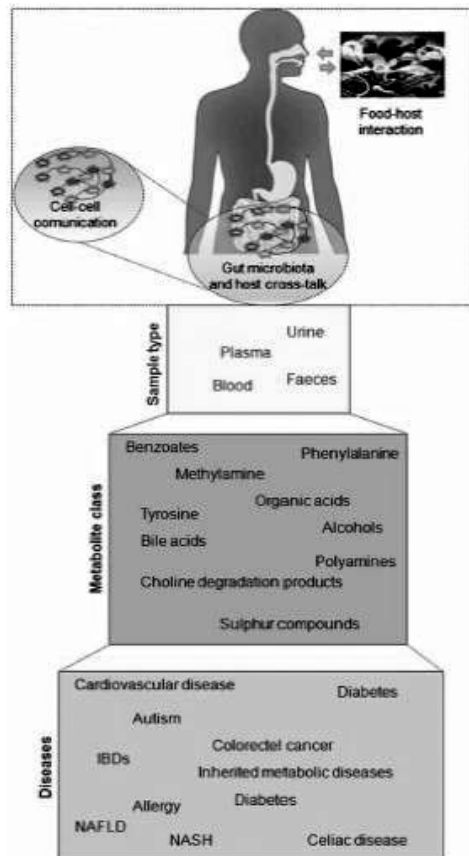


Fig. 1. Schematic representation of diet, microbes, and host interaction at gut level. The chemical dialogue via low molecular weight metabolites, peptides, and proteins between cell-cell and host-microbes leads to the metabolite production in different body fluids which could be considered as disease biomarkers.

분변 중의 대사체를 연구하여 Crohn병, 위궤양, Hirschsprung병, 소아지방변증, 알러지 등 염증성 장질환(Inflammatory bowel disease, IBDs)에 대한 진단 정보를 찾을 수 있었다. Metabolomics는 잠재력있는 응용 학문이다. 이 분야의 접목으로 기능성 유전체 연구, 시스템 생물학, 약학, 독성학, 영양 유전학이 가능해졌다. 발견대사계수(discovery metabolite profiling, DMP)는 proteome과 metabolome 사이를 연계하는 특정 효소에 의해 생성된 모든 대사물질을 분석한다.

대사물질 분석 기술

Metabolomics는 복잡한 세포 개체의 다변량 대사 반응 연구를 포함하는 대사체계 전반을 연구하는 학문이다. 이상적으로는 소변, 혈액, 기타 체액 채취, 기기분석, 질병이나 약물 부작용을 감지하는 화학물질 프로파일 확인과 같은 전개로 연구를 진행한다. 그러나 여기엔 문제점이 있다. 한 사람의 대사체 프로파일이 다른 사람에게도 일반화될 수 없다는 것인데, 생활습관, 질병, 식이, 영양상태가 변수로 작용한다.

수백개의 대사체를 분리·동정하는 것은 특정 질병 유발과 관련이 있는 생활습관, 식이가 인체에 주는 영향에 대해 많은 정보를 제공한다. GC, LC, HPLC, UPLC, MS, CE, FTIR, NMR 등의 최신 기술의 빠른 발달로 분리, 검색, 대사체 정량이 가능해졌고, 대사 경로 또한 확인할 수 있게 되었다. Metabolomics는 시스템을 이루는 구성 요소의 복잡한 상호작용이나 개체의 개별 특성 보다는 세포의 항상성을 확인할 수 있는 전반적인 시스템을 알아내는데 초점을 맞추고 있다. 설사, NMR, GC-MS, LC-MS를 이용한다 하더라도, 모든 대사체

를 완벽하게 밝혀내는 기술이란 것은 없다. 그러나, MS를 기반으로 한 metabolomics는 대사체의 정량과 정성에 높은 선별도와 감수성을 제공하고, 고효율 물질 분리 기술은 대사물질 분리의 번거로움을 줄여준다(Table 1).

NMR 기반 metabolomics는 특정 상황에서 생산되는 대사체에 대한 전체적인 이해에 도움이 된다. 특히 체액에 대한 비파괴적 대사체 연구와 장내 미생물 또는 건강 장애 관련 식이 영향 연구에 유용하다.

보통 장내 미생물 대사 분석은 분변 시료를 대상으로 한다. 분변에 대한 $^1\text{H-NMR}$ 모니터링은 위궤양, IBD 등의 진단 정보를 제공한다. Saric 등(2008)은 $^1\text{H-NMR}$ 을 이용해 사람, 생쥐, 쥐 등 다른 포유류의 유사점과 차이점을 연구한 바 있다.

소화기 대사체의 주요 발단은 주로 인체와 미생물이 처리하는 식이 영양원이다. Martin 등(2012)은 분변의 $^1\text{H-NMR}$ 분석을 통해 probiotics의 공급이 장내 미생물총 상호작용에 큰 영향을 준다고 발표하였다. Probiotics의 공급은 *Lactobacillus casei*의 단백질 대사 변화에 영향을 주었다.

Dumas 등(2006)은 $^1\text{H-NMR}$ 기술을 이용하여 식이변화로 유도된 장내 미생물과 인체의 co-metabotype 간의 복잡한 연관성을 연구하기 위해 실험 동물과 장내 미생물의 유전자간 특징을 연구하였다. Ndagijimana 등은 $^1\text{H-NMR}$ 로 건강한 사람의 분변을 분석하여 *Lactobacillus acidophilus*, *Bifidobacterium longum*과 fructooligosaccharides로 된 synbiotic food 공급의 효과를 연구하였다. 그 결과, 상당수 또는 상당 범위의 대사물질이 synbiotic food 섭취에 영향을 받았으며, 독특한 대사적 특성이 있음을 보고하였다. Jacobs 등은 사람 분변의 물/에탄올 추출물의 $^1\text{H-NMR}$ 대사 프로파일을 조사하였다. $^1\text{H-NMR}$ 은 소변이

나 plasma 분석에 많이 이용하는 기술 중 하나이다. 그 첫 연구는 콩을 포함한 식이를 섭취했을 때, 소변 내 대사체를 본 것이다. 소변 시료에서 hippurate와 glycine 같은 주요 대사 성분을 확인하는 것으로 카모마일차, 커피, 와인, 차의 소화반응을 연구할 수 있었다.

심혈관 질환이 의심되는 환자가 운동을 해서 생길 수 있는 국소성 빈혈(Ischemia)의 발생 위험은 ¹H-NMR을 이용한 혈액분석으로 미리 예견할 수 있다. Lactate, glucose, 지질, 장쇄 지방산이 여기에 포함된 주요 대사산물이다. Xanthine과 ascorbate는 동맥 경화 동물 모델에서 plaque 형성 마커로 여겨지고, lipoprotein subclasses는

¹H-NMR을 이용한 방법으로 분석해낼 수 있다. 그러나, ¹H-NMR로 매우 가치있는 정보 분석이 가능하다 하더라도 낮은 해상도와 감도로 인해 사용에는 제한적일 수밖에 없다.

GC-MS 기반 metabolomics는 고효율 시료 처리, 표준 retention time과 mass spectra로 정확한 peak 분리가 가능하다. GC-MS는 metabolomics에 광범위하게 사용되고, 효과적이고, 재현성 있는 분석이 가능하다. 유기산, 아미노산, 당알콜, 방향족 아민, 지방산 프로파일 또한 얻을 수 있다. GC-MS/SPME(Solid Phase MicroExtraction) 기반 분석법은 분변 시료를 정량·정성 분석할 수 있는 빠르고도 매우 효과적인 분석법이

Table 1. Common analytical techniques used in metabolomics

분석 방법	장점	단점	비고
NMR	<ul style="list-style-type: none"> · 빠른 분석 · 우수한 분석력 · 유도체 불필요 · 비파괴 	<ul style="list-style-type: none"> · 낮은 감도 · 복잡한 스펙트럼 · 구성요소 하나당 한 개 이상의 peak · 복잡한 matrix로 인한 라이브러리 제한성 	<ul style="list-style-type: none"> · 화학적 고려사항: 분리한 대사물질에 특히 2D-NMR을 사용하면 상세한 구조정보를 얻을 수 있음. · 화학적 이용 제한: 이 방법은 화학적 제한점이 거의 없어 시료에 바로 적용 가능 · 속도: 수 분에서 수 시간. 자성의 강도에 따라 감도가 증가할 수 있음.
GC-MS	<ul style="list-style-type: none"> · 민감 · 확실성 · 광범위한 선형 범위 · 다량의 상업화된 도는 공공 라이브러리 	<ul style="list-style-type: none"> · 느림 · 유도체 필요 경우 많음. · 분석물 대다수가 온도에 불안정하며 분석하기에 큼. 	<ul style="list-style-type: none"> · 화학적 고려사항: 단독 사용으로는 대사체 동정이 어려움. 그러나, MS와 NMR과 함께 사용하면 매우 유용한 수단이 됨 · 화학적 이용 제한: 용매 추출이 필요. 유도체 필요. · 속도: 분리에 매우 유용. 보통 10~30분 소요.
LC-MS	<ul style="list-style-type: none"> · (대개의 경우) 유도체 불필요 · 다양한 분리 가능 · 많은 샘플 용량 	<ul style="list-style-type: none"> · 느림 · 상업적 libraries 제한적 	<ul style="list-style-type: none"> · 화학적 고려사항: 단독 사용으로는 대사체 동정이 어려움, 그러나, MS와 NMR과 함께 사용하면 매우 유용한 수단이 됨 · 화학적 이용 제한: 극성 물질에 더 유용 · 속도: 분리에 매우 유용. 보통 10~30분 소요.
FT-IR	<ul style="list-style-type: none"> · 빠른 분석 · 시료의 화학적 조성을 완벽하게 확인 · 유도체 불필요 	<ul style="list-style-type: none"> · 매우 난해한 스펙트럼 · 한 개 요소에 한 개 이상의 peak · 대사산물 동정은 거의 불가능 · 시료건조 과정 필요 	<ul style="list-style-type: none"> · 화학적 고려사항: 제한적인 구조 정보 제공. 기능기 동정에 유용 · 화학적 이용 제한: 시료를 바로 측정 가능. · 속도: 10~60초

다. 특히, GC-MS/SPME는 생체 시료의 대사체 프로파일을 연구하는 우수한 방법으로 이 기술을 이용하여 신생아와 성인의 분변을 비교하였고, 소화기 질환의 특징적인 휘발성 마커를 규명하였다. GC-MS/SPME로는 주로 함황 화합물, 질소 화합물, pyrazines, indoles, aldehyde, ketone, esters, 알콜, 페놀, 유기산, 탄화수소와 furan, furanones 같은 스트레스 물질을 확인할 수 있다. Vitali 등은 이 방법을 사용하여 synbiotic food가 사람 장내 미생물의 생태와 대사체 프로파일에 미치는 영향을 관찰한 바 있다. 정상인에선 큰 차이가 없던 장내 미생물이 synbiotic food를 먹은 후, 장내 대사 활성에 큰 변화가 생겼다. 분변 대사체 프로파일에 대한 정규 판별법(Canonical discrimination Analysis of Principle coordinates, CAP)은 synbiotic food 섭취에 따른 개체 차이를 보여준다. 이처럼 최근 GC-MS/SPME는 사람 장내 미생물에 대한 prebiotics 또는 probiotics 영향을 연구하는 데 많이 사용하고 있다.

2011년 Zheng 등은 GC-MS를 이용하여 정상 Wistar rats의 미생물 대사산물 profile과 광범위 β -lactam계 항생제 처리 동물의 미생물 대사산물 profile을 비교 연구하였다. 이 연구를 통해 이전에는 보고되지 않은 oligopeptides, 당질 등 소변에서 202종, 분변에서 223종의 대사체를 분리·동정하였다. Maccaferri 등은 rifaximin 투여가 미생물 대사 profile에 미치는 영향을 GC-MS/SPME를 이용하여 연구한 바 있다. 이 기술은 $^1\text{H-NMR}$ 기술과 병용하여 gluten-free diet를 받는 19명의 환아들과 15명의 건강한 아동의 소변 및 분변 시료의 metabolome에 대해 비교 연구하였다. 이 연구로 celiac disease에 의해 휘발성 유기 화합물과 유리 아미노산 농도에 큰 변화가 있다는 것이 밝혀졌다. 또, Francavilla는 GC-MS/

SPME를 이용하여 우유 알러지가 있는 유아에서 lactose가 장내 미생물총 조성과의 생성되는 대사물질에 미치는 영향에 대하여 조사 보고한 바 있다.

MS와 HPLC 기술은 화합물 특성을 구조수준까지 알려주는데 많이 이용된다. Metabolomics 분야에서는 MS와 HPLC를 모두 이용하여 복잡한 생체 시스템을 거치면서 발생하는 알려지지 않은 내인적/외인적 특성을 밝히는데 사용한다. LC를 이용하는 것은 광범위한 극성을 가진 각각의 물질을 최소한의 전처리 작업만으로도 분리할 수 있는 가장 유용한 분리 방법일 것이다. 실제로 LC-MS를 기반으로 하는 metabolomic 기술은 인체 질병 진단을 위한 이용이 증가하고 있다. GC-MS와 LC-MS를 함께 사용하여 영양부족 동물모델(rats)의 특징적인 대사체 발굴에도 이용된 바 있다.

NMR처럼 진동형 분광기술인 RAMAN과 FTIR과 감도면에서 견줄만하다. 그러나 FTIR은 처리효율이 높은 스크리닝 방법이자 생체 시료 분류방법으로 꼽힌다. 특히 살아있는 세포에서 생화학적 변화를 살펴보기 위한 FTIR 분광기의 사용은 최근 10년동안 많은 중요 정보를 얻을 수 있게 해주었다. 이 기술로 서로 다른 파장대의 적외선을 흡수하는 대사물질을 *in vivo* / *in vitro* 조건에서 분석하는 것이 가능해졌다. 분자나 시스템의 생체 관련성은 FTIR 분석으로 생체막, 동물 조직, 미생물 세포 배양, 임상 시료에 더하여 지질, 지방산, 단백질, peptides, 당질, 핵산같은 분자 분석까지도 가능해졌다. 이 방법은 특히 단백질이나 지질같은 분리된 생체분자를 분석하는데 사용하고, 특히 최근에는 자궁경부암, 파킨슨병, 알츠하이머병, 결석, 관절염 같은 복잡한 시료분석을 지원한다. FTIR 분광기술은 생체에서 구조나 구성에 관한 다양한 정보를 분자 수준까지 효과적으로 제시한다. Li 등은 FTIR 분광기술과 chemometrics를 이용하여 조직

시료를 분석하여 colitis와 대장암을 확인해낼 수 있다고 보고하였다. 또한, Argov 등은 FTIR 현미 분광계법으로 유사한 증상의 환자에서 IBDs와 대장암을 구별해내기도 하였다. 이렇듯 FTIR 분광기술은 질병의 진단에 확실하고, 신속하고 효과적이며 경제적인 방법으로 활용되고 있다.

지금까지 설명한 기술 중 어느 것을 택할 것인가는 분석하고자하는 목표에 달려 있다. (1) HPLC, GC-MS, LC-MS는 알려진 대사물질이나 특정 대사 경로의 생성물의 정성이나 정량 분석에 사용한다. (2) HPLC-MS, CE-MS, LC-NMR은 대사물질 profiling에 사용하고, (3) NMR, DIMS (Direct Infusion Electrospray Ionization-MS), LDI-MS(Laser Desorption Ionization-MS), MSLDI-MS(Matrix Suppressed-MS), FTIR, Raman 분광법은 차이점 규명을 위한 대사체 pro-

file 비교 등과 같은 감식 작업에 사용한다(Fig 2).

또한 Chromatography-MS 시스템은 metabolomics에서 대사체를 검출분리 하는데 가장 유리한 방법으로 여겨진다. GC-TOF(Time of Flight)-MS, 이차원 GC coupled to TOF MS(GCxGC-TOF-MS), HPLC-MS, 분석에 큰 장점이 있는 UPLC-MS와 CE-MS는 대사체 profiling이나 특정 목적이 있는 분석적 접근법 등 포유류의 대사체 연구에 많이 사용된다. GC-MS 나 GC-MS/SPME는 대사체 연구에 광범위하게 사용되는데 특히, GC-MS/SPME는 어려운 전처리 과정이 반드시 필요하지 않다. Garner 등은 괴사성 장염증세를 보이는 미숙아와 정상 유아의 분변에서 휘발성 화합물 profile이 다르다는 것을 GC-MS/SPME를 이용한 연구로 보여줬다. 최근의 연구들은 건강한 사람이나 환자들에 대한 연구

APPROACH	TARGETED ANALYSIS	METABOLITE PROFILING	METABOLIC FINGERPRINTING
STRENGTH	<ul style="list-style-type: none"> •Quantitative •Low detection limit •High throughput 	Global image of the sample	<ul style="list-style-type: none"> •Global image of the sample •Directly applicable to pattern recognition •Highest throughput
PITFALL	<ul style="list-style-type: none"> •Limited number of targeted compounds •No detection of untargeted compounds •Need for the pure targeted compounds for the calibration •Expensive 	<ul style="list-style-type: none"> •Semi-quantitative •Difficulties in informatics •Medium throughput 	No compound identification
TECHNIQUES	HPLC, GC-MS, LC-MS, NMR	HPLC-MS, CE-MS, LC-NMR	NMR, DIMS, LDHMS, MSLDI-MS, FTIR, RAMAN

Fig. 2. Different approaches and respective techniques: pitfalls and strengths. The employed technique depends on the followed approach: targeted analysis, metabolite profiling, and metabolic fingerprinting.

에서 장내 미생물 또한 중요한 생화학적 기능을 수행한다는 것을 밝혀내고 있다.

대사체 분석 자료에서 가치있는 결론을 추출하는 것은 매우 중요하다. 여기에는 서로 다른 소프트웨어를 이용한 다양한 방법이 있을 수 있다. 일부 개인차로 나타나는 생물학적인 다양성, 시료 채취, 시료 처리, 분석 방법과 같은 일부 parameters는 결과의 재현성에 영향을 미치기 때문에 반복 실험과 분석을 수행해야만 한다. 특히 생물학적인 차이는 분석차보다 더 큰 것이 일반적이다. 또, NMR과 FTIR 분광법외에도 GC, 특히 LC에서 retention time의 변화는 흔하게 나타난다. NMR에서 재현성이 나타나지 않는 것은 기기적인 이유라기 보단 시료 준비와 큰 관련성이 있다. 그러나 엄격하게 조건을 통제한다 해도 차이는 발생할 수 있다. 이러한 이유로 NMR은 고해상도(HiRes)가 의미있는 dataset을 만드는데 적합한 반면, MS 응용 기술에는 MetAlign, XCMS, MZmine 같은 signal 정렬 소프트웨어의 사용이 반드시 포함된다.

Datasets의 통합 관리법

특정 마커를 질병, 영양상태, 또는 식이법과 연관시키는 것은 어렵다. 대신에 시스템 전체를 조사하고 동시에 여러 경로를 탐색하여 건강 또는 병적 상태에 관한 더 큰 지표를 얻을 수 있다. 지금까지 수백 종의 대사물질은 높은 효율로 처리해내는 새로운 기술들은 많은 데이터를 양산했고, 데이터의 견고성, 실용성, 적용성, 분석 효율성을 갖추기 위해 수년간 하드웨어 면에서도 소프트웨어 면에서도 꾸준한 업그레이드가 진행되어 왔다. 그러므로, 다변량 통계분석 도구와 같은 강력한 분석전략의 결합이 새로운 “black box”를 발견하기 위해, 복

잡한 생물학적 시스템에 대한 최대의 관련 정보와 지식을 추출하기 위해 필요하다.

단일 또는 병행 분석 기술의 사용 또는 다른 종류의 시료를 분석하는 작업은 수집한 수천종의 데이터를 어떻게 관리하고 표현할 것인가라는 큰 문제에 필연적으로 직면하게 된다. 시스템 생물학자들의 희망사항 중 하나는 이 데이터들이 분자수준에서 네트워크를 이뤄 성장, 노화, 건강 및 질병 상태에 대한 통찰력을 가진 전체적인 그림이 그려지도록 통합되는 것이다. 적절한 실험 설계, 샘플 숫자, 통계 분석은 정확하고 유효한 가설 및 생물학적 결론을 도출하기 위해 필요하다. 강력한 다변량 분석 기법은 데이터를 얻고, 구조관련 생체 시스템의 유사성이나 차이점을 규명하는데 필요하다.

데이터 단순화를 위한 첫 번째 단계는 정해진 방법에 따라 한 개 peak의 상대적인 강도를 색깔로 표현되는 블록으로 대체한 Heat Maps를 사용하는 것이다. 이 방법은 처음 microarray 연구를 위해 개발 되었으나, 대사체 연구에도 사용되고 있다. Rajaram과 Oono는 2010년 그들의 연구에서 전통적인 Heat Maps를 대체할 주요소 분석법(principal component analysis, PCA)과 다차원 스케일링(multi-dimensional scaling, MDS)을 결합한 2·3차원 신규 플롯 NeatMap을 제안하였다. 비록 NeatMap이 지금껏 유전체학 연구에서만 사용되었으나, 대사체학에서도 역시 사용될 수 있다. 두 번째 단계는 서로 다른 datasets을 해석할 수 있도록 다변량 통계 분석 방법을 사용하는 것이다. Richards 등(2010)은 자료 통합을 위한 세가지 전략을 제시하였다. (1) Metabonomics, genomics, proteomics, transcriptomics 처럼 다른 ‘-omic’ 기술 간 또는 ‘-omic’ 플랫폼에서 얻은 자료 통합, (2) 플랫폼 간 또는 서로 다른 분광 플랫폼(NMR, GC-MS, LC-MS 등)으로부터 데이

터의 통합, (3) 샘플 간, 또는 다른 사람의 샘플 (혈장, 뇨, 조직, 대변)에서 얻은 데이터의 통합이 그것이다(Fig 3).

가장 널리 사용되는 다변량 분석법은 PCA와 PLS(Partial Least Squares)법이다. 특히, PCA는 NMR 또는 GC-MS분석 profiling데이터의 다변량 분석에 사용한다. PLS 및 그로부터 파생된 OPLS(Orthogonal PLS)도 NMR 데이터를 profiling하는 데 사용되었다. PLS와 비교하면, OPLS는 문제를 판별하는데 있어 더 높은 수준의 해법을 유도해 해석하기 쉽고 분명한 모델을 만들어 낼 수 있다. PCA와 PLS를 잠재적으로 연계된

여러 가지 다른 “block”이라고 불리는 datasets를 해석하는데 사용할 때, spectrum에서의 연계변이(co-variants)로 인해 로딩 플롯(loading plots)은 일반적으로 복잡하고, 해당 스코어 플롯(score plot)에 상관관계를 갖는 것도 어렵다. 이런 이유에서 멀티 블록 기법이 사용되어야 한다.

계층적 다중 블록 분할 기술(HPCA, HPLS)은 서브 스펙트럼으로 스펙트럼을 블록화하고 PCA에 의해 서브 스펙트럼을 투영하여 원래의 데이터로부터 생성된 새로운 변수를 기준으로 한다. 이 새로운 변수는 예기치 않은 빛의 산란이나 무작위 signals을 감소시키고, signals를 체계적인 정보를

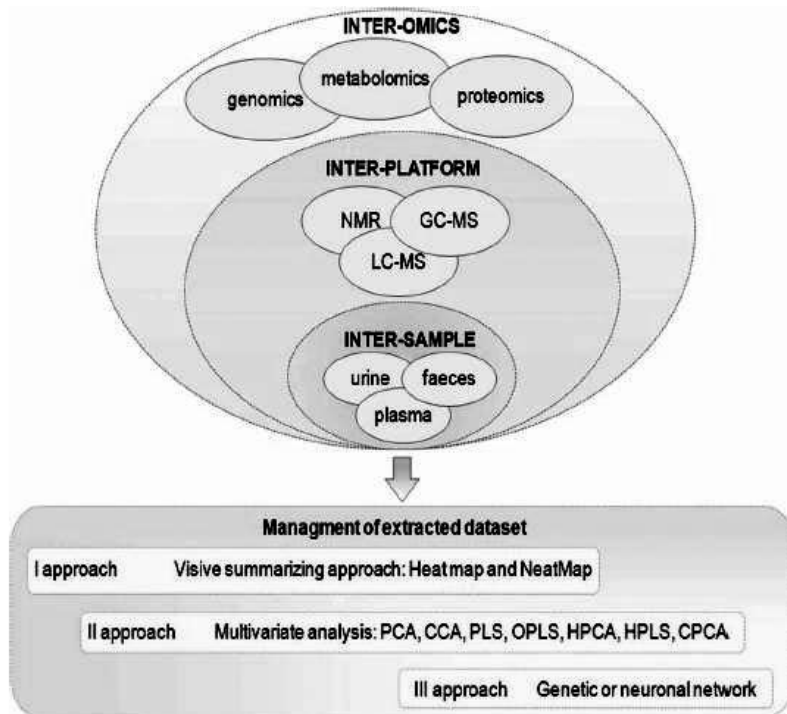


Fig. 3. Schematic representation of statistical data integration methods in the area of inter-omic, inter-platform, and inter-sample integration. The management of different dataset derived from the metabolomic approaches are integrated by the most diffuse multivariate methods.

유지하기 위해 이 후의 PCA 또는 PLS 계산에 사용된다. 이 기술은 scores와 로딩(loading) 사이의 상관 관계를 쉽게 해석한다는 가장 큰 장점이 있다.

사용한 멀티 블럭 PCA의 또 다른 종류는 Consensus PCA(CPCA)로, 가능한 많은 변수를 설명할 수 있는 경향을 찾아 특정 대사산물과 그 외 대사산물간의 관계를 규명하는데 초점을 맞춘 데이터 분석이 가능하다. CPCA는 같은 대상을 측정한 서술자변수(descriptor variables)의 여러 블록을 비교하는 방법으로 소개되었고, HPCA에 대한 미미한 차이는 데이터 표준화 되었다. 또한, 이러한 정준 상관 분석(Canonical Correlation Analysis, CCA) 및 regularized CCA 등의 다변량 분석은 다른 다양한 채널에서 얻어진 datasets를 통합하는데 광범위하게 사용된다.

생물학적 시스템의 이해에 대한 전체적인 접근 방식이 지속적으로 발전함에 따라, datasets를 분석하고 통합하는 데 사용되는 기술은 발생하는 문제를 관리하기 위해 지속적으로 기술을 업데이트해야한다. 그렇기 때문에 datasets 관리로 얻어진

결과가 지닌 의미를 해석하는 시스템 생물학자들의 역할은 매우 중요하다.

● 자료출처 ●

P.Vernocchi, L.Vannini, D.Gottardi, F.Del Chierico, D.I.Serrazanetti, M.Ndagijimana and M.E.Guerzoni, Integration of datasets from different analytical techniques to assess the impact of nutrition on human metabolome. *Frontiers in Cellular and Infection Microbiology*, 2, 1-10, 2012

이 영 경 이학박사

소 속 : 한국식품연구원 공정기술연구단

전문분야 : 발효 및 기능성 미생물학

E-mail : ykrhee@kfri.re.kr

T E L : 031-780-9319