바이오연료 연구와 Proteomics - Part 2

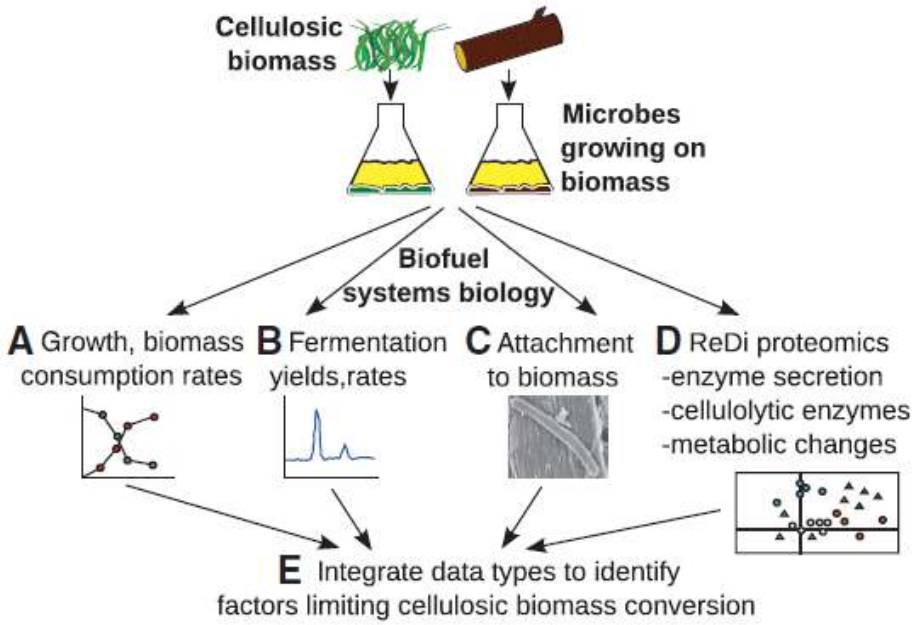
박정진

바이오연료 중에서 2세대 연료로 가장 가까운 미래에 상업화가 가능한 것으로 알려져 있는 셀룰로스는 사실 세계에서 가장 풍부한 바이오연료 자원 중 하나이다. 하지만, 이러한 셀룰로스는 준결정구조 (quasicrystalline structure)를 갖는 분자량이 큰 다당류 (polysaccharide)로 이루어져 있어, 이러한 구조를 해체하는 과정이 셀룰로스 유래 바이오연료를 만드는데 있어 가장 중요한 과정으로 알려져 있다.

이를 해결하기 위해, 요즘 많은 연구자 들에게 각광을 받고 있는 방법은 *Clostridium phytofermentans*와 같은 미생물을 이용하는 것이다. 이러한 미생물 들은 바이오매스의 구조를 해체하는 것 뿐만이 아니라, 오탄당이나 육탄당을 에탄올로 변환시키는 역할을 한다. 그래서 이러한 대사과정과 관련되어 있는 효소를 최적화하는 연구가 많이 진행되고 있는 중이다.

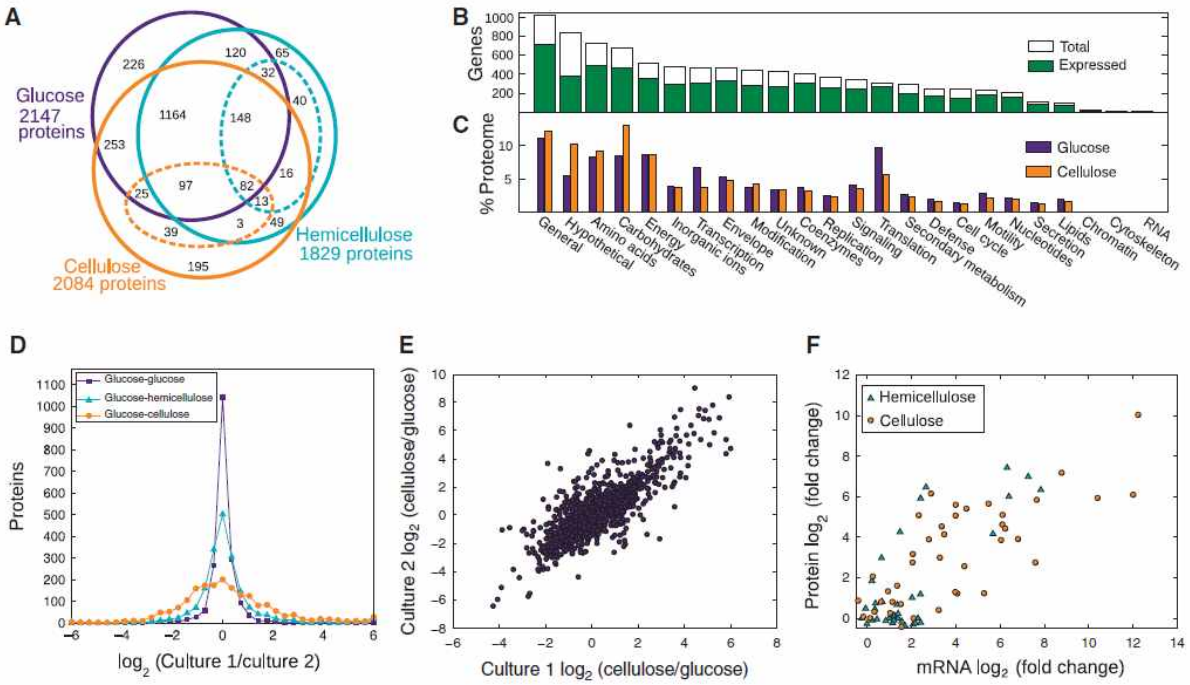
하지만 이러한 과정에 연관되어 있는 효소는 매우 복잡한 대사과정과 얽혀있어, 사람들이 원하는 대사산물을 최대한도로 생산해 내기 위해서는, systems-level 분석이 필요하다. 즉, 단순히 몇개의 mRNA나 단백질 발현 정도만을 보는 것이 아니라 최대한 많은 수의 다양한 카테고리에 속한 (metabolite, protein, 그리고 mRNA) 데이터를 확보하여 그것들을 통합적으로 분석하는 방법이 필요하다.

최근 미국 하바드 의대와 Northeastern 대학교에서 발표한 논문[1]은 이러한 바이오연료 연구에서 proteomics가 어떠한 역할을 맡을 수 있을 지 보여주는 한 예라고 할 수 있을 것 같아, 그들의 연구 진행 방법을 소개해 보고자 한다.



<그림 1> 셀룰로스 유래 에탄올 연구를 위한 systems-level 연구 방법

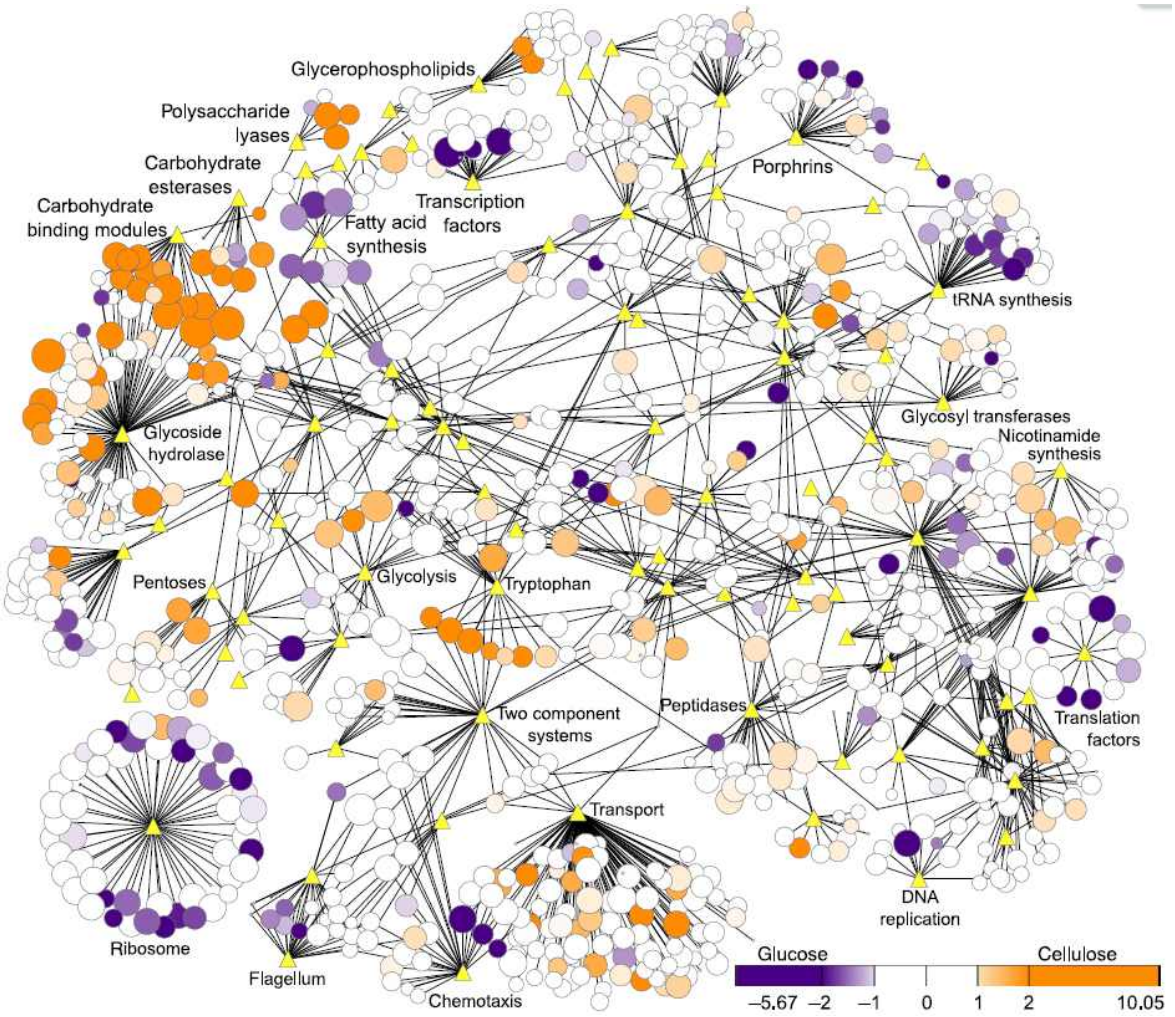
<그림 1>에서 보면, 일반적인 실험군과 대조군을 정한 후, A 에서 C 와 같이 전통적인 배양 데이터를 확보한다. 그리고 D 에서와 같이 단백질 전체에 대한 정량 비교와 mRNA 발현량 비교, 그리고 대표적인 대사산물의 생산량 변화를 관찰한 다음, E 에서와 같이 복합적인 분석을 하는 것이다.이를 위해 이 연구진은 앞서 말한 정량 분석 방법 중에서, 단백질 전체 정량에 적합한 APEX 방법[2]을 이용했다.



<그림 2> 질량분석기를 이용한 단백질 정성분석과 (A 에서 C 까지) 정량분석 (D 에서 F 까지) 결과 (A) 각각 다른 처리에 따라 확인된 단백질 갯수, 전체 3926 개의 단백질 중에 putative protein 이 2567 (65%)를 차지하고 있음 (B) COG (Clusters of Orthologous Genes)의 분류에 따른 유전자 중에서 65% 가량의 단백질이 매칭되었음 (C) 각기 다른 탄소원(glucose 와 cellulose)에서 배양된 미생물에서 추출된 단백질의 정량분석 결과의 합 (D) 두번에 걸친 반복 실험에서 단백질 발현량 비교 결과 (E) 두번에 걸친 실험에서 단백질 fold-change 결과 (F) 탄수화물 대사 과정에 있는 40 개의 단백질과 유전자의 발현량 비교

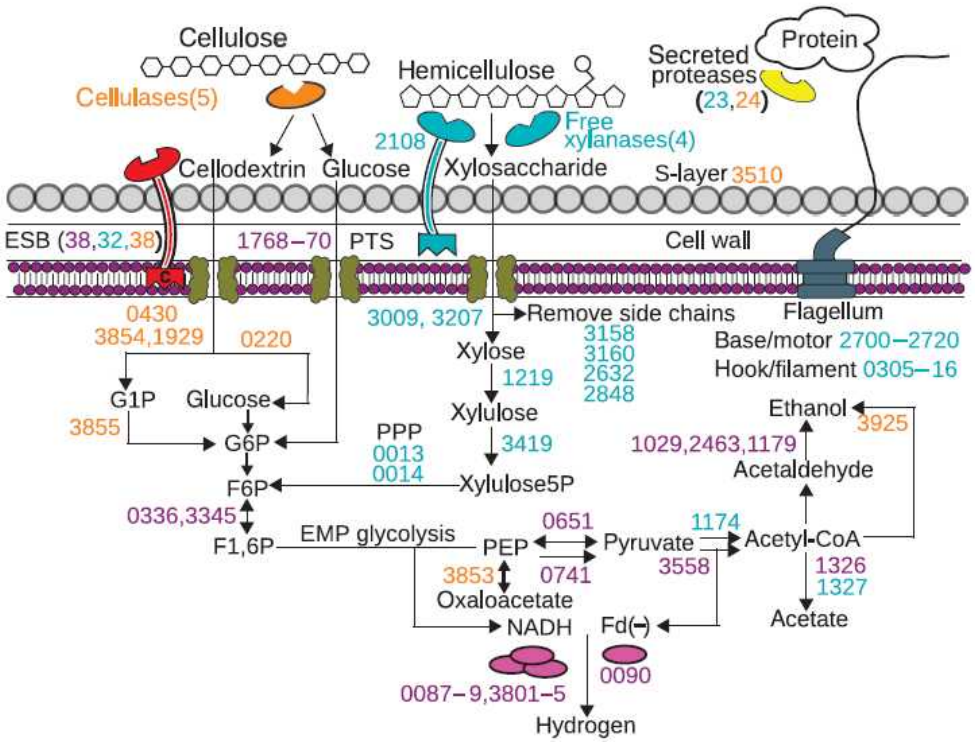
본 실험의 proteomics 결과는 <그림 2>에 나온 바와 같이, 우선 2 천여개의 단백질이 각 샘플마다 검출되었으며 그 중 65% 가량은 putative protein 이었다. 이를 가지고 우선

COGs (http://www.ncbi.nlm.nih.gov/COG/)에 따른 분류를 수행하였으며, 검출된 단백질 중 65%가 이 분류를 따랐다. 이를 가지고 우선 재현성을 확인해 보았으며 (<그림 2>의 D 와 E) 탄수화물 대사와 관련이 있는 40 개의 mRNA 발현량과 비교해 보았다. 이를 통해 단백질과 선별된 mRNA 의 발현량 사이에는 상당히 높은 상관 관계가 있다는 것을 보여 주었다 (r^2 값이 0.7 근처).



<그림 3> Cytoscape[3] 를 이용한 단백질체 발현량 비교 - 동그란 교점 (node)은 각각의 단백질을 나타내고 있으며 세모 모양의 노란색 교점은 KEGG/carbohydrate-active enzyme (CAZy) 카테고리를 나타내고 있다. 그리고 선은 KEGG 나 CAZy 에 의해 정의된 단백질 간의 관계를 나타낸다. Glucose 와 cellolose 를 첨가한 실험에서 단백질 발현량의 차이가 2 배보다 작은 경우에는 흰색을 사용했으며, glucose 에서 2 배보다 많이 발현되면 보라색을, cellolose 를 첨가했을 때 2 배 이상 많이 발현된 경우에는 오렌지색을 사용

<그림 3>에서는 생물학적 대사 과정이 아니라, 순전히 단백질 간의 상관 관계나 카테고리만을 사용하여 발현량에 차이가 나는 단백질들을 나타내 보았다. 사용된 프로그램은 cytoscape (http://www.cytoscape.org/) 였으며, 이를 통해 선입견 없이 (이미 알려져 있는 대사 과정) 중요한 효소일 수 있다고 판단되는 단백질 들을 찾을 수 있었으며, 이를 가지고 <그림 4>와 같은 모델을 제시할 수 있었다.



<그림 4> 식물 바이오매스를 분해하고 발효시키는데 관여하는 중요 단백질 모델. 나타낸 숫자는 그 기작에 관여하는 것으로 보이는 단백질의 번호

Proteomics 란 유전자의 명령으로 만들어진 단백질체를 대상으로 유전자의 기능, 단백질의 기능 이상 및 구조변형 유무 등을 규명하고 추적하는 기술이다. 이 기술이 각광을 받는 이유는 유전체 염기서열이 다 밝혀졌다고 하더라도 (Genomics) 그것만 가지고는 유전자 산물의 기능을 알 수가 없고 (Transcriptomics), 이것이 translation 되어 단백질이 만들어 지는데, 이 단백질을 분석하지 않고서는 그 유전자의 세포 내 기능을 알 수가 없기 때문이다. 이 기술의 장점은 이처럼 명백하지만, 지금까지 이 기술을 바이오연료 분야에 제대로 적용시킨 예는 드물었었다. 이 논문에 나온 것처럼 proteomics 를 이용한다면 셀룰로스 유래 에탄올 연구의 중요 부분을 밝혀 내는 것 뿐만

아니라 미생물을 이용한 유용 물질 생산과 관련된 연구 분야 전체에도 적용이 가능할 것으로 예상된다.

1 Tolonen, A. C. *et al.* Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol Syst Biol* **7**, 461, doi:http://www.nature.com/msb/journal/v7/n1/suppinfo/msb2010116_S1.html (2011).

2 Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotech* **25**, 117-124, doi:http://www.nature.com/nbt/journal/v25/n1/suppinfo/nbt1270_S1.html (2007).

3 Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).