

## Whole-transcript expression analysis 기술 소개

(주)서린바이오사이언스 최혁진 Hyuk-Jin Choi(aatc@seoulin.co.kr)

전통적으로, microarray는 각 유전자의 3' 말단에 있는 수백 base 부위에서 probe를 design하여 전체 유전자의 발현치로 사용하여 왔다. 이러한 접근은 3' oligo(dT)로 priming하고 labeling하는 실험 방법으로 이용되었으며 전체 유전자 발현을 관찰하는데 적절하다고 생각되어왔다. 그러나, 이것은 각 유전자의 3' 말단이 확실하게 밝혀지고, 각 transcript는 손상되지 않은 poly-A를 가지고, 전체 유전자는 한 단위로 발현된다는 가정에서 출발한다.

하지만, 이러한 가정은 모든 유전자 또는 모든 샘플에 적용할 수 없다. 알려져 있는 유전자의 60% 이상이 alternatively splice되며, 별개의 기능을 가지는 잠재적인 수 십만 개의 transcript variant가 존재한다. 50% 만큼의 질병연관 point mutation이 splice pattern의 변화에 의한 결과로 나타날 수 있고, 20%의 암 유발 돌연변이가 exon-skipping 기전에 따른 결과로 나타날 수 있다. 아쉽게도 전통적인 3' expression microarray들은 동일한 3' 말단을 가지는 alternative spliced transcript들을 구분하지 못한다. Alternative splicing에 의해 3' exon이 결실된 transcript, non-polyadenylation, 유전체 결실, 중복 등은 3' expression microarray로 확인할 수 없다.

### "See the REAL biology"

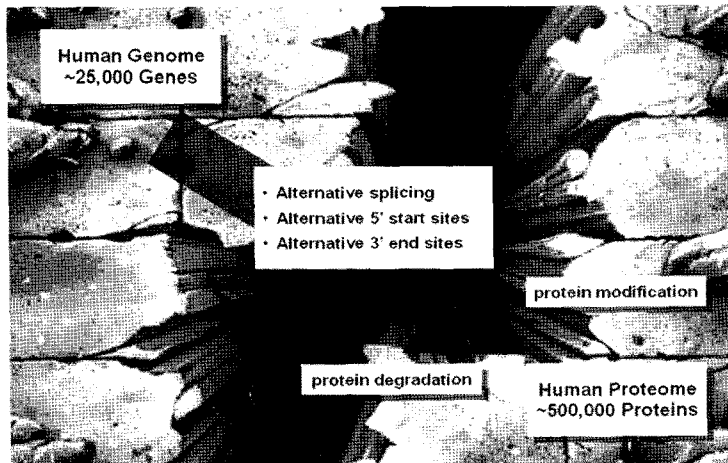
따라서, Gene level이 아닌 transcript level에서 어떤 일이 발생하는지를 확인하는 것이 필요하게 된다. 그림 1에서

와 같이 Human gene의 수와 Human proteom의 수에는 큰 차이가 존재한다.

이러한 차이가 발생하는 이유는 protein modification, protein degradation등에 의한 것도 있지만, 상당수가 alternative splicing, alternative 5' start site, alternative 3' end site등에 기인한 것이다. 그러므로 이러한 차이를 파악하기 위해서는 exon-level의 발현을 관찰할 필요가 있다.

"전통적인 microarray들은 이미 반복적으로 실험을 통해 증명되어져 왔기 때문에, 위의 결과는 Exon array의 성능이 probe set의 mapping 뿐만 아니라 전체 transcriptome을 더욱 자세히 관찰하기 위해 추가된 수천 개의 probe set에 대해서도 역시 신뢰할만하다는 증거이다."

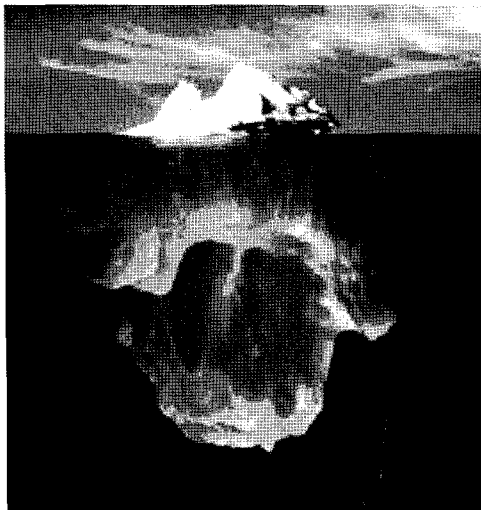
Okoniewski M., et al.



[그림 1] Human Genome과 Human Proteom간의 차이와 그 이유

GeneChip\_ Exon 1.0 ST Array와 GeneChip\_ Gene 1.0 ST Array와 같은 Affymetrix의 Whole Transcript(WT) Assay와 고밀도 어레이들은 전체적인 유전자 발현에 대해 보다 완벽하고 정확한 그림을 제공하여 연구자들이 존재하는지 알지 못하던 transcript isoform들을 검출할 수 있게 해주고 이전에는 동일하게 보이던 transcript들을 구별할 수 있게 해준다.

그림 2와 같이 3' expression microarray는 유전체 병산의 일각만을 보여주지만, Whole Transcript(WT) Assay는 병산 전체를 보여주는 것이다.



【그림 2】 3' expression array로는 유전체 병산의 일각만을 볼 수 있다.

### Human, Mouse, Rat Exon 1.0 ST Array

Exon당 약 4개의 probe, 유전자당 약 40개의 probe들을 가지고 있는 GeneChip\_ Exon 1.0 ST Array는 한번의 실험으로 두 가지 형태- Exon 단위, Gene 단위-의 발현 데이터를 제공한다.

Exon별로 있는 여러 probe들의 Exon-level 분석은 alternative splicing과 각 exon별 발현 등 가장 높은 해상도의 microarray 분석을 진행할 수 있다. Gene-level로는 발현 수준뿐만 아니라 정확하게 무엇이 발현되는지를 파악할 수 있다. 이러한 특징들은 질병 감수성과 병리학 연구

### 질병연구에서 Alternative splicing의 중요성

“최근의 연구에서 alternative splice 산물의 70%~88%가 단백질로 전환되는 것이 나타났다. 이러한 변환의 대부분이 기능적으로 매우 흥미로운 것으로 나타났다”

Modrek B, and Lee C.

등에서 매우 중요한 역할을 할 수 있다.

Gene-level 분석에서, 다른 exon들에 있는 여러 probe들은 한 유전자로부터 나오는 모든 transcript들에 대한 단일 발현데이터로 요약된다.

Exon array는 모든 microarray의 디자인 중 가장 넓은 coverage를 제공하는데, 이는 연구자가 알려져 있지 않은 splicing 산물도 찾아낼 수 있게 하기 위함이다. 또한, Exon array의 probe 디자인은 Uni-Gene cluster가 아닌 유전체 염기서열로부터 진행된 것이므로 서열의 annotation을 손쉽게 갱신할 수 있다. 각 probe들은 genome의 서열과 연계되므로 다양한 유전체 정보들-SNP, 돌연변이, 염색체 결손과 증가 등-과 통합하여 분석을 하는 것이 가능하게 된다.

### GeneChip\_ Human Gene 1.0 ST Array

Human Gene 1.0 ST Array는 역시 Whole-transcript assay에 의해 실험이 진행되는 신제품으로, 유전자당 약 26개의 probe들이 들어있다. 3' 말단으로 치우쳐진 expression array 디자인에 비해 보다 정확하고 보다 완벽한 그림을 제공할 수 있다. Human Gene 1.0 ST Array는 Exon 1.0 ST Array에서 Gene 단위 발현을 보기 위해 부분적으로 선택된 probe들로 제작된 array이다.

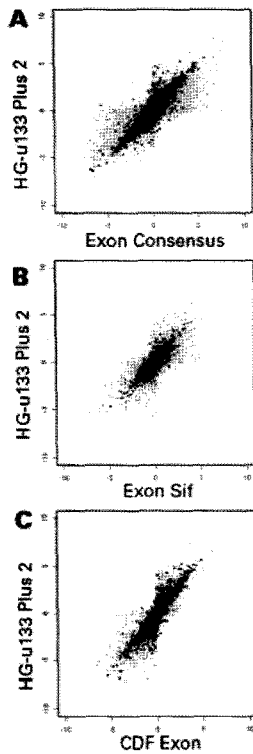
### Alternative splicing 현상의 대두

“서로 다른 그룹에서 진행된 많은 연구들에서 40%~60%에 해당하는 인간 유전자의 상당부분에 alternative splice 형태가 발견되고 있다.”

Lee C, and Roy M.

Whole-transcript Array의 결과들이 3' Expression microarray 결과로 검증되었다.

최근에 진행된 연구에 따르면 Whole-transcript array가 검증된 Affymetrix의 3' gene expression microarray 결과와 일치되는 것이 보고되었다. Okoniewski는 최근에 Exon 1.0 ST Array와 3' 기반 디자인의 Human Genome U133 plus 2.0의 결과를 비교하였는데, 높은 일치율을 보였다(그림 3).



**[그림 3]** Affymetrix exon과 일반적인 expression array간의 높은 일치도를 보여주는 scatter plot.

### Exon-level 분석의 실제

보다 정확한 유전자 발현분석을 위해, exon array는 전체 유전자에서 나타나는 발현 변화를 밝힐 수 있는 유일한 방법이다. 다음의 사례와 같이 연구자들은 이미 질병진행과정에서 중요한 역할을 할 수 있는 exon의 독특한 변화를 밝혀내고 있다.

- 질병의 진행상황에 따른 비정상적인 splicing event의 발견 - Drs. Pim French and Justine Peeters,

Erasmus Medical Center, Netherlands.

- Splicing 조절 기전의 발견 - Bruno, et al., University of Texas M. D. Anderson Cancer Center
- 유전된 SNP들과 splicing 패턴과의 연관성 - Dr. Jacek Majewski, McGill University and Genome Quebec Innovation Center
- Exon-level integration of proteomics and microarray data - Bitton D. A., BMC Bioinformatics 9:118(2008)
- Rapid high-resolution karyotyping with precise identification of chromosome breakpoints - Mao X., Genes, Chromosomes and Cancer 46(7):675-83(2007)
- Genome-wide analysis of antisense transcription with Affymetrix exon array - GE X., BMS Genomics 9:27 (2008)
- A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity - Huang R. S., PNAS 104(23):9758-9763 (2007)

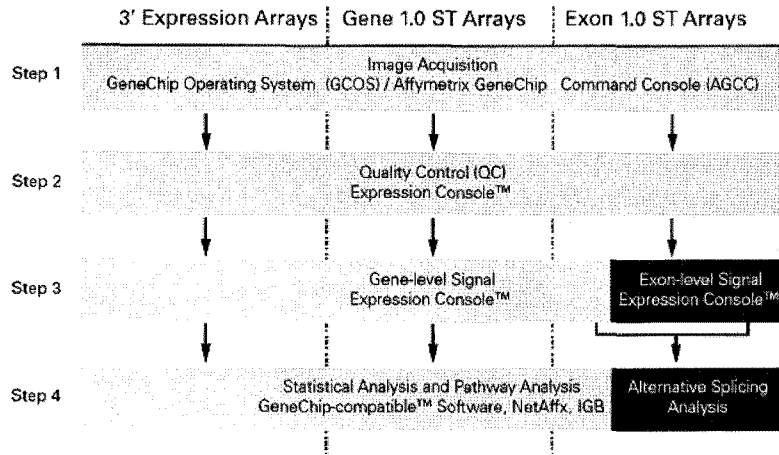
### 데이터 분석 솔루션

#### Gene-level analysis

Exon 1.0 ST, Gene 1.0 ST array에서 gene-level 발현을 분석하는 것은 3' expression array를 분석하는 것과 유사하므로 간단하다.

먼저, GeneChip® Operating System(GCOS) 또는 Affymetrix GeneChip® Command Console(AGCC)에서 초기의 image를 얻는다. 그리고 Affymetrix Expression Console software의 사용하기 용이한 workflow에 따라 quality control(QC)와 probe set 요약을 진행하여 gene-level의 signal data를 얻는다(그림 4).

Affymetrix의 partner인 GeneChip-compatible software 제조사들은 유익하게 발현 조절된 유전자들의 목록을 얻는데 필요한 통계적 분석방법을 제공한다. 또한, 출판된 문헌과 데이터베이스를 통해 유전자간의 상호작용을



[ 그림 4 ] Gene 1.0 ST와 Exon 1.0 ST array의 Data 분석 workflow. 3' expression array와 유사하며, GCOS 또는 Affymetrix GeneChip Command Console(AGCC), Expression Console, GeneChip-compatible software, NetAffx Analysis Center 와 IGB등을 사용하여 진행한다.

	Biotique's Xray	Genomatix Genomatix ChipInspector	imp JMP Genomics	Partek Partek® Genomix Suite™
Supports analysis of GeneChip Human Gene 1.0 ST Array	Yes	Yes	Yes	Yes
Supports analysis of GeneChip Exon Arrays	Yes	Yes	Yes	Yes
Supports analysis for 3' expression arrays	No	Yes	Yes	Yes
Other GeneChip-compatible offerings	—	Regulation (ChIP-on-chip)	SNP Analysis	<ul style="list-style-type: none"> <li>Tiling (Regulation, ChIP-on-chip)</li> <li>SNP Analysis</li> <li>Chromosomal Copy Number &amp; LOH Analysis</li> </ul>
Application Highlights	<ul style="list-style-type: none"> <li>Fast MS Excel 5.5Mn offers familiar Excel interface</li> <li>Rigorous normalization and analysis for unlimited numbers of CEL files on computers with 500 MB ram</li> <li>Dynamically generated methods, results and array quality document suitable for publications</li> </ul>	<ul style="list-style-type: none"> <li>Analysis based on single probe</li> <li>Single probe creation based on latest genome annotation</li> <li>Currently 37 arrays 128 genomes supported, including promoter tiling arrays</li> <li>Tight integration with other bioinformatics tools to derive greater biological insight</li> </ul>	<ul style="list-style-type: none"> <li>Powerful SAS analysis enhanced by interactive JMP graphics</li> <li>Streamlined, automated experimental design tools</li> <li>Quality control (QC) tools including batch effect, reference</li> <li>Enhance predictive modeling capabilities</li> </ul>	<ul style="list-style-type: none"> <li>Fast and memory-efficient</li> <li>Comprehensive statistics and interactive visualization</li> <li>Retrieves batch effects</li> <li>Fast-turn workflow for tiling, copy number and tiling arrays</li> <li>Integration of results from multiple GeneChip technologies such as expression, regulation and copy number</li> <li>Built and vetted diagnostic and prognostic classifiers</li> </ul>
Website	www.biotique.com	www.genomatix.de	www.imp.com/genomics	www.partek.com

[ 표 ] 유전자 발현과 alternative splicing 분석을 지원하는 GeneChip - compatible 제품군

시각화 시켜주는 pathway analysis를 위한 다양한 software 제조사들이 있다(표).

NetAffx Analysis Center는 Affymetrix에서 무료로 제공하며, 연구자들이 어레이 디자인과 annotation정보들을 연관시켜 볼 수 있도록 해주는 체계이다.

Integrated Genome Browser(IGB)는 다양한 데이터소스로부터 유전체와 annotation을 연결시켜 시각화 시켜주는 도구이다.

### Exon-level analysis

Expression Console은 연구자들이 exon array 데이터에서 exon-level의 signal을 계산할 수 있게 해준다. Alternative transcript를 예측하기 위해서는 GeneChip-compatible software와 Exon 1.0 ST array로 gene-level과 exon-level signal의 통계분석을 진행한다. Pathway analysis는 NetAffx Analysis Center와 IGB에서 얻은 annotation으로 exon array data와 분석을 진행할 수 있는데, 3' expression array data와 비슷한 workflow로 진행한다.