



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0086197
(43) 공개일자 2016년07월19일

(51) 국제특허분류(Int. Cl.)
G01N 33/68 (2006.01) G01N 30/72 (2006.01)
(52) CPC특허분류
G01N 33/6848 (2013.01)
G01N 30/72 (2013.01)
(21) 출원번호 10-2015-0003625
(22) 출원일자 2015년01월09일
심사청구일자 2015년01월09일

(71) 출원인
한양대학교 산학협력단
서울특별시 성동구 왕십리로 222(행당동, 한양대학교내)
승실대학교산학협력단
서울특별시 동작구 상도로 369 (상도동)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
백은옥
서울특별시 서초구 효령로 164, 1동 103호 (방배동, 신동아아파트)
이상원
서울특별시 성북구 인촌로17가길 64, 108동 602호 (안암동1가, 래미안안암아파트)
(74) 대리인
특허법인 무한

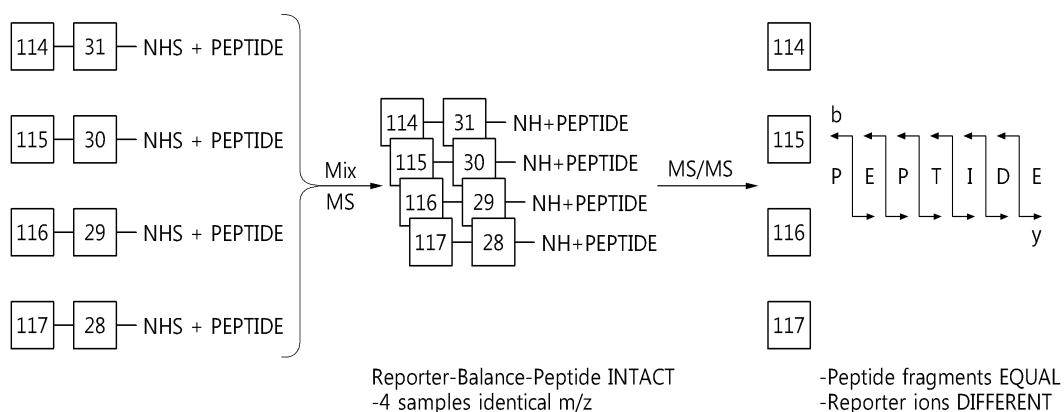
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 동중 원소 태그에 기초한 정량적 질량 분석의 잡음을 고려한 펩타이드의 동정 및 정량 방법 및 시스템

(57) 요약

동중 원소 태그에 기초한 정량 질량 분석의 잡음을 고려한 펩타이드의 동정 및 정량 방법 및 분석 시스템이 개시된다. iTRAQ와 같이 동중 원소 태그 기반의 정량 방법은 MS/MS 단계에서 함께 단편화(cofragmentation)됨으로써 펩타이드의 정량 및 동정 결과의 오류가 발생할 수 있다. 이와 같이, 함께 단편화된 효과를 확인하기 위해, 분석 시스템은 미리 알려진 전구체 구성과 다양한 전구체 분리 순도를 가지는 혼합 스펙트럼을 시뮬레이션할 수 있다.

대표도 - 도2



(72) 발명자

황규백

서울특별시 송파구 올림픽로 99, 153동 1203호 (잠실동, 잠실엘스)

이홍란

서울특별시 동작구 상도로 369, 정보과학관 409호 (상도동, 숭실대학교)

김호근

서울특별시 성북구 솔샘로 120, 124동 103호 (정릉동, 정릉풍림아이원아파트)

문동기

서울특별시 성북구 아리랑로 89, 102동 1604호 (돈암동, 일신건영휴먼빌아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호 2012M3A9B9036676

부처명 미래창조과학부

연구관리전문기관 (재)한국연구재단

연구사업명 원천기술개발사업/바이오의료기술개발사업/다중오믹스신기술개발사업

연구과제명 차세대시퀀싱 기반 프로테옴 데이터 해석을 위한 소프트웨어 기술

기 여 율 1/1

주관기관 한양대학교 산학협력단

연구기간 2014.06.01 ~ 2015.05.31

명세서

청구범위

청구항 1

동중 원소 태그를 이용한 펩타이드의 정량 자료를 이용하여 잡음이 포함되지 않은 제1 스펙트럼을 추출하는 단계;

상기 제1 스펙트럼에 대해 펩타이드 동정을 수행하는 단계;

상기 추출된 제1 스펙트럼을 혼합하여 잡음이 포함된 제2 스펙트럼을 생성하는 단계;

상기 제2 스펙트럼에 대해 펩타이드 동정을 수행하는 단계;

상기 제1 스펙트럼 및 제2 스펙트럼 각각에 매칭되는 펩타이드들을 추출하여 비교하는 단계;

상기 제1 스펙트럼 및 제2 스펙트럼에 대해 펩타이드 정량을 수행하는 단계;

상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과를 비교하는 단계;

상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과 간의 비교 결과에 따라 펩타이드의 발현 비율에 따른 제1종 오류 확률 또는 제2종 오류 확률을 제어할 수 있는 발현 차이 펩타이드(Differentially-Expressed Peptide: DEP)를 검출하기 위한 임계값을 결정하는 단계

를 포함하는 펩타이드의 동정 및 정량 방법.

청구항 2

제1항에 있어서,

상기 제1 스펙트럼을 추출하는 단계는,

적어도 하나의 펩타이드 동정 도구를 이용하여 동중 원소 태그를 통해 도출된 펩타이드의 스펙트럼을 동정하는 단계;

전구체 분리 순도(Precursor Isolation Purity: PIP)에 기초하여 상기 동정을 통해 확인된 펩타이드들 중 잡음이 포함되지 않은 펩타이드들을 선택하는 단계; 및

상기 선택된 펩타이드들의 전하량 및 리포터 이온을 고려하여 펩타이드 동정 도구를 통해 동정된 제1 스펙트럼을 추출하는 단계

를 포함하는 펩타이드의 동정 및 정량 방법.

청구항 3

제1항에 있어서,

상기 제2 스펙트럼을 생성하는 단계는,

전구체의 질량대 전하비가 분리 윈도우(isolation window)에 존재하는 2개의 제1 스펙트럼을 랜덤하게 선택하는 단계;

상기 랜덤하게 선택된 2개의 제1 스펙트럼들을 가공하는 단계;

제1 스펙트럼들을 미리 설정된 전구체 분리 순도에 대응하는 혼합 비율에 따라 상기 가공된 2개의 제1 스펙트럼들을 혼합하는 단계

를 포함하는 펩타이드의 동정 및 정량 방법.

청구항 4

제3항에 있어서,

상기 2개의 제1 스펙트럼을 가공하는 단계는,

상기 2개의 제1스펙트럼의 전체 이온 카운트(Total Ion Count)에 기초하여 상기 2개의 제1스펙트럼을 정규화하는 단계;

상기 정규화된 2개의 제1스펙트럼의 피크들을 혼합하는 단계를 포함하는 펩타이드의 동정 및 정량 방법.

청구항 5

제4항에 있어서,

상기 피크들을 혼합하는 단계는,

리포터 이온의 피크를 비닝하는 단계; 및

상기 리포터 이온을 제외한 나머지 펩타이드 피크를 병합하는 단계

를 포함하는 펩타이드의 동정 및 정량 방법.

청구항 6

제5항에 있어서,

상기 비닝하는 단계는,

상기 리포터 이온들의 이론적인 질량을 기준으로 리포터 이온에 대해 미리 설정된 윈도우 내에 존재하는 리포터 이온의 피크를 하나의 피크로 병합하는 펩타이드의 동정 및 정량 방법.

청구항 7

제5항에 있어서,

상기 병합하는 단계는,

상기 나머지 펩타이드 피크의 질량대전하비에 따라 두 개의 정규화된 제1 스펙트럼의 피크들을 병합하는 펩타이드의 동정 및 정량 방법.

청구항 8

제1항에 있어서,

상기 펩타이드 동정을 수행하는 단계는,

MODa, MS-GF+, Proteome Discoverer 중 적어도 하나를 포함하는 펩타이드 동정 도구를 이용하여 펩타이드 동정을 수행하는 펩타이드의 동정 및 정량 방법.

청구항 9

제1항에 있어서,

상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과를 비교하는 단계는,

상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과의 차이를 확률 분포로 출력하는 펩타이드의 동정 및 정량 방법.

청구항 10

제1항에 있어서,

상기 임계값을 결정하는 단계는,

제1 스펙트럼의 발현 차이, 제2 스펙트럼의 발현 차이 및 제1 스펙트럼의 DEP를 검출하기 위한 임계값을 이용하여 제2 스펙트럼의 DEP를 검출하기 위한 임계값을 결정하는 펩타이드의 동정 및 정량 방법.

청구항 11

분석 시스템에 있어서,

동중 원소 태그를 이용한 펩타이드의 정량 자료를 이용하여 잡음이 포함되지 않은 제1 스펙트럼을 추출하는 제1 스펙트럼 추출부;

상기 제1 스펙트럼에 대해 펩타이드 동정을 수행하는 제1 펩타이드 동정부;

상기 추출된 제1 스펙트럼을 혼합하여 잡음이 포함된 제2 스펙트럼을 생성하는 제2 스펙트럼 추출부;

상기 제2 스펙트럼에 대해 펩타이드 동정을 수행하는 제2 펩타이드 동정부

상기 제1 스펙트럼 및 제2 스펙트럼 각각에 매칭되는 펩타이드들을 추출하여 비교하는 펩타이드 비교부;

상기 제1 스펙트럼 및 제2 스펙트럼에 대해 펩타이드 정량을 수행하는 펩타이드 정량부;

상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과를 비교하는 정량 결과 비교부;

상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과 간의 비교 결과에 따라 펩타이드의 발현 비율에 따른 제1종 오류 확률 또는 제2종 오류 확률을 제어할 수 있는 발현 차이 펩타이드(Differentially-Expressed Peptide: DEP)를 검출하기 위한 임계값을 결정하는 임계값 결정부

를 포함하는 분석 시스템.

청구항 12

제11항에 있어서,

상기 제1 스펙트럼 추출부는,

적어도 하나의 펩타이드 동정 도구를 이용하여 동중 원소 태그를 통해 도출된 펩타이드의 스펙트럼을 동정하고,

전구체 분리 순도(Precursor Isolation Purity: PIP)에 기초하여 상기 동정을 통해 확인된 펩타이드들 중 잡음이 포함되지 않은 펩타이드들을 선택하며,

상기 선택된 펩타이드들의 전하량 및 리포터 이온을 고려하여 펩타이드 동정 도구를 통해 동정된 제1 스펙트럼을 추출하는 분석 시스템.

청구항 13

제11항에 있어서,

상기 제2 스펙트럼 생성부는,

전구체의 질량대 전하비가 분리 윈도우(isolation window)에 존재하는 2개의 제1 스펙트럼을 랜덤하게 선택하고,

상기 랜덤하게 선택된 2개의 제1 스펙트럼들을 가공하며,

제1 스펙트럼들을 미리 설정된 전구체 분리 순도에 대응하는 혼합 비율에 따라 상기 가공된 2개의 제1 스펙트럼들을 혼합하는 분석 시스템.

청구항 14

제11항에 있어서,

상기 제2 스펙트럼 생성부는,

상기 2개의 제1스펙트럼의 전체 이온 카운트(Total Ion Count)에 기초하여 상기 2개의 제1스펙트럼을 정규화하고, 상기 정규화된 2개의 제1스펙트럼의 피크들을 혼합하는 분석 시스템.

청구항 15

제14항에 있어서,

상기 제2 스펙트럼 생성부는,

리포터 이온의 피크를 비닝하고, 상기 리포터 이온을 제외한 나머지 펩타이드 피크를 병합하는 분석 시스템.

청구항 16

제15항에 있어서,

상기 제2 스펙트럼 생성부는,

상기 리포터 이온들의 이론적인 질량을 기준으로 리포터 이온에 대해 미리 설정된 윈도우 내에 존재하는 리포터 이온의 피크를 하나의 피크로 병합함으로써 리포터 이온의 피크를 비닝하는 분석 시스템.

청구항 17

제15항에 있어서,

상기 제2 스펙트럼 생성부는,

상기 나머지 펩타이드 피크의 질량대전하비에 따라 두 개의 정규화된 제1 스펙트럼의 피크들을 병합함으로써 펩타이드 피크를 병합하는 분석 시스템.

청구항 18

제11항에 있어서,

상기 펩타이드 동정부는,

MODa, MS-GF+, Proteome Discoverer 중 적어도 하나를 포함하는 펩타이드 동정 도구를 이용하여 펩타이드 동정을 수행하는 분석 시스템.

청구항 19

제11항에 있어서,

상기 정량 결과 비교부는,

상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과의 차이를 확률 분포로 출력하는 분석 시스템.

청구항 20

제11항에 있어서,

상기 임계값 결정부는,

제1 스펙트럼의 발현 차이, 제2 스펙트럼의 발현 차이 및 제1 스펙트럼의 DEP를 검출하기 위한 임계값을 이용하여 제2 스펙트럼의 DEP를 검출하기 위한 임계값을 결정하는 분석 시스템.

발명의 설명

기술 분야

[0001] 이하의 실시예들은, 펩타이드의 동정 및 정량 방법 및 시스템에 관한 것으로, 보다 구체적으로는 동중 원소 태그에 기초한 정량적 질량 분석의 잡음을 고려한 펩타이드의 동정 및 정량 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 질량 분석(mass spectrometry)에 기초한 단백질의 정량(quantification)은 전구체(precursor) 내에서 다수의 표본(sample)에 존재하는 펩타이드의 상대적인 정량을 확인할 수 있는 방법을 의미한다. 구체적으로, 특정 전구체 내에 둘 이상의 표본에서 추출된 동일한 서열을 가지는 펩타이드(peptide)가 포함될 수 있다. 그러면, 질량 분석을 통해 단백질을 정량함으로써, 서로 다른 표본들 각각에 포함된 동일한 서열을 가진 펩타이드 간의 상대적인 정량 정보가 확인될 수 있다.

[0003] 질량 분석에 기초한 단백질의 정량은 생명 과학 연구, 임상 연구를 비롯한 다양한 분야에서 광범위하게 사용되고 있다. 이 때, 단백질의 정량은 동위 원소 또는 동중 원소를 표지로 이용하는 기법과 비표지 기법으로 구분될

수 있다.

[0004] iTRAQ 및 TMT와 같은 동중 원소 태그 기반의 정량은 질량 분석 기반의 정량을 위해 많이 활용되고 있다. 동중 원소 태그를 이용하는 단백질의 정량은 동위 원소를 이용하는 방법에 비해서 표본의 준비가 용이한 장점이 있다. 그리고, 동중 원소 태그를 이용하는 단백질의 정량은 정량의 비교 대상이 되는 펩타이드들이 동일한 스펙트럼에 존재하기 때문에 비표지 기법에 비해 상대적으로 정확한 정량을 수행할 수 있다는 장점이 있다.

[0005] 하지만, 위와 같은 장점에도 불구하고, 동중 원소 태그를 이용하는 단백질의 정량은 유사한 질량대 전하비(mass-to-charge ratio)를 나타내는 서로 다른 펩타이드들의 전구체가 MS/MS(tandem mass spectrometry)로 인해서 함께 분리(co-isolation)될 수 있다. 이러한 분리로 인해 함께 단편화(co-fragmentation)됨으로써 왜곡된 리포터 이온 피크들(reporter ion peaks)이 생성되고, 최종적으로는 잘못된 펩타이드의 정량 정보가 도출될 수 있다. 다시 말해서, 동중 원소 태그 기반의 정량은 복수의 펩타이드들이 함께 단편화(co-fragmentation)되기 때문에, 타겟 펩타이드의 동정 및 정량 결과가 부정확한 문제가 있다. 이러한 문제로 인해서, 암 환자의 표본과 정상인의 표본 사이의 발현 차이 펩타이드가(differentially expressed peptide) 잘못 추정될 수 있다.

[0006] 이와 같은 펩타이드의 정량 정보의 왜곡으로 인해서, 서로 다른 조건에서의 발현 차이 펩타이드들이 잘못 검출될 수 있다. 펩타이드의 정량 정보의 왜곡은 (i) 실제로는 펩타이드들 간의 정량의 차이가 존재하지 않지만, 잡음으로 인해서 정량의 차이가 존재하는 것으로 왜곡되는 비율 확장(ratio inflation), 또는 (ii) 실제로 펩타이드들 간의 정량의 차이가 존재하지만 잡음으로 인해서 정량의 차이가 존재하지 않는 것으로 왜곡되는 비율 축소(ratio compression)를 야기시킨다.

[0007] 다시 말해서, 비율 확장은 펩타이드들 간의 원래의 상대적인 정량비보다 잡음으로 인한 상대적인 정량비가 더 큰 경우를 의미한다. 그리고, 비율 축소는 펩타이드들 간의 원래의 상대적인 정량비보다 잡음으로 인한 상대적인 정량비가 더 작은 경우를 의미한다.

[0008] 따라서, 동중 원소를 태그로 이용하는 질량 분석에 기초한 단백질의 정량에 있어서, 잡음으로 인한 정량 정보의 왜곡 문제를 해결하기 위한 방안이 필요하다.

발명의 내용

해결하려는 과제

[0009] 본 발명은 동일한 전구체 내에 포함된 펩타이드에 대한 잡음의 정도에 따른 왜곡의 정도를 파악하기 위해 모의 자료를 생성하는 정량 방법을 이용한다.

[0010] 본 발명은 모의 자료를 기반으로 펩타이드에 대한 잡음의 정도에 따라 펩타이드의 실제 정량이 왜곡됨에 따른 비율 확장 또는 비율 축소에 의한 위양성(false positive) 및 위음성(false negative) 발현차이 펩타이드(differentially-expressed peptide) 검출 확률을 제어할 수 있는 임계값을 도출하는 정량 방법을 이용한다.

과제의 해결 수단

[0011] 본 발명의 일실시예에 따른 펩타이드의 동정 및 정량 방법은, 동중 원소 태그를 이용한 펩타이드의 정량 자료를 이용하여 잡음이 포함되지 않은 제1 스펙트럼을 추출하는 단계; 상기 제1 스펙트럼에 대해 펩타이드 동정을 수행하는 단계; 상기 추출된 제1 스펙트럼을 혼합하여 잡음이 포함된 제2 스펙트럼을 생성하는 단계; 상기 제2 스펙트럼에 대해 펩타이드 동정을 수행하는 단계; 상기 제1 스펙트럼 및 제2 스펙트럼 각각에 매칭되는 펩타이드들을 추출하여 비교하는 단계; 상기 제1 스펙트럼 및 제2 스펙트럼에 대해 펩타이드 정량을 수행하는 단계; 상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과를 비교하는 단계; 상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과 간의 비교 결과에 따라 펩타이드의 발현 비율에 따른 제1종 오류 확률 또는 제2종 오류 확률을 제어할 수 있는 발현 차이 펩타이드(Differentially-Expressed Peptide: DEP)를 검출하기 위한 임계값을 결정하는 단계를 포함할 수 있다.

[0012] 상기 제1 스펙트럼을 추출하는 단계는, 적어도 하나의 펩타이드 동정 도구를 이용하여 동중 원소 태그를 통해 도출된 펩타이드의 스펙트럼을 동정하는 단계; 전구체 분리 순도(Precursor Isolation Purity: PIP)에 기초하여 상기 동정을 통해 확인된 펩타이드들 중 잡음이 포함되지 않은 펩타이드들을 선택하는 단계; 및 상기 선택된 펩타이드들의 전하량 및 리포터 이온을 고려하여 펩타이드 동정 도구를 통해 동정된 제1 스펙트럼을 추출하는 단계를 포함할 수 있다.

- [0013] 상기 제2 스펙트럼을 생성하는 단계는, 전구체의 질량대 전하비가 분리 윈도우(isolation window)에 존재하는 2개의 제1 스펙트럼을 랜덤하게 선택하는 단계; 상기 랜덤하게 선택된 2개의 제1 스펙트럼들을 가공하는 단계; 제1 스펙트럼들을 미리 설정된 전구체 분리 순도에 대응하는 혼합 비율에 따라 상기 가공된 2개의 제1 스펙트럼들을 혼합하는 단계를 포함할 수 있다.
- [0014] 상기 2개의 제1 스펙트럼을 가공하는 단계는, 상기 2개의 제1스펙트럼의 전체 이온 카운트(Total Ion Count)에 기초하여 상기 2개의 제1스펙트럼을 정규화하는 단계; 및 상기 정규화된 2개의 제1스펙트럼의 피크들을 혼합하는 단계를 포함할 수 있다.
- [0015] 상기 혼합하는 단계는, 리포터 이온의 피크를 비닝하는 단계; 및 상기 리포터 이온을 제외한 나머지 펩타이드 피크를 병합하는 단계를 포함할 수 있다.
- [0016] 상기 비닝하는 단계는, 상기 리포터 이온들의 이론적인 질량을 기준으로 리포터 이온에 대해 미리 설정된 윈도우 내에 존재하는 리포터 이온의 피크를 하나의 피크로 병합할 수 있다.
- [0017] 상기 병합하는 단계는, 상기 나머지 펩타이드 피크의 질량대전하비에 따라 두 개의 정규화된 제1 스펙트럼의 피크들을 병합할 수 있다.
- [0018] 상기 펩타이드 동정을 수행하는 단계는, MODa, MS-GF+, Proteome Discoverer 중 적어도 하나를 포함하는 펩타이드 동정 도구를 이용하여 펩타이드 동정을 수행할 수 있다.
- [0019] 상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과를 비교하는 단계는, 상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과의 차이를 확률 분포로 출력할 수 있다.
- [0020] 상기 임계값을 결정하는 단계는, 제1 스펙트럼의 발현 차이, 제2 스펙트럼의 발현 차이 및 제1 스펙트럼의 DEP를 검출하기 위한 임계값을 이용하여 제2 스펙트럼의 DEP를 검출하기 위한 임계값을 결정할 수 있다.
- [0021] 본 발명의 일실시예에 따른 분석 시스템은 동중 원소 태그를 이용한 펩타이드의 정량 자료를 이용하여 잡음이 포함되지 않은 제1 스펙트럼을 추출하는 제1 스펙트럼 추출부; 상기 제1 스펙트럼에 대해 펩타이드 동정을 수행하는 제1 펩타이드 동정부; 상기 추출된 제1 스펙트럼을 혼합하여 잡음이 포함된 제2 스펙트럼을 생성하는 제2 스펙트럼 추출부; 상기 제2 스펙트럼에 대해 펩타이드 동정을 수행하는 제2 펩타이드 동정부; 상기 제1 스펙트럼 및 제2 스펙트럼 각각에 매칭되는 펩타이드들을 추출하여 비교하는 펩타이드 비교부; 상기 제1 스펙트럼 및 제2 스펙트럼에 대해 펩타이드 정량을 수행하는 펩타이드 정량부; 상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과를 비교하는 정량 결과 비교부; 상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과 간의 비교 결과에 따라 펩타이드의 발현 비율에 따른 제1종 오류 확률 또는 제2종 오류 확률을 제어할 수 있는 발현 차이 펩타이드(Differentially-Expressed Peptide: DEP)를 검출하기 위한 임계값을 결정하는 임계값 결정부를 포함할 수 있다.
- [0022] 상기 제1 스펙트럼 추출부는, 적어도 하나의 펩타이드 동정 도구를 이용하여 동중 원소 태그를 통해 도출된 펩타이드의 스펙트럼을 동정하고, 전구체 분리 순도(Precursor Isolation Purity: PIP)에 기초하여 상기 동정을 통해 확인된 펩타이드들 중 잡음이 포함되지 않은 펩타이드들을 선택하며, 상기 선택된 펩타이드들의 전하량 및 리포터 이온을 고려하여 펩타이드 동정 도구를 통해 동정된 제1 스펙트럼을 추출할 수 있다.
- [0023] 상기 제2 스펙트럼 생성부는, 전구체의 질량대 전하비가 분리 윈도우(isolation window)에 존재하는 2개의 제1 스펙트럼을 랜덤하게 선택하고, 상기 랜덤하게 선택된 2개의 제1 스펙트럼들을 가공하며, 제1 스펙트럼들을 미리 설정된 전구체 분리 순도에 대응하는 혼합 비율에 따라 상기 가공된 2개의 제1 스펙트럼들을 혼합할 수 있다.
- [0024] 상기 제2 스펙트럼 생성부는, 상기 2개의 제1스펙트럼의 전체 이온 카운트(Total Ion Count)에 기초하여 상기 2개의 제1스펙트럼을 정규화하고, 상기 정규화된 2개의 제1스펙트럼의 피크들을 혼합할 수 있다.
- [0025] 상기 제2 스펙트럼 생성부는, 리포터 이온의 피크를 비닝하고, 상기 리포터 이온을 제외한 나머지 펩타이드 피크를 병합할 수 있다.
- [0026] 상기 제2 스펙트럼 생성부는, 상기 리포터 이온들의 이론적인 질량을 기준으로 리포터 이온에 대해 미리 설정된 윈도우 내에 존재하는 리포터 이온의 피크를 하나의 피크로 병합함으로써 리포터 이온의 피크를 비닝할 수 있다.
- [0027] 상기 제2 스펙트럼 생성부는, 상기 나머지 펩타이드 피크의 질량대전하비에 따라 두 개의 정규화된 제1 스펙트

럼의 피크들을 병합함으로써 펩타이드 피크를 병합할 수 있다.

- [0028] 상기 펩타이드 동정부는, MODa, MS-GF+, Proteome Discoverer 중 적어도 하나를 포함하는 펩타이드 동정 도구를 이용하여 펩타이드 동정을 수행할 수 있다.
- [0029] 상기 정량 결과 비교부는, 상기 제1 스펙트럼의 정량 결과와 제2 스펙트럼의 정량 결과의 차이를 확률 분포로 출력할 수 있다.
- [0030] 상기 임계값 결정부는, 제1 스펙트럼의 발현 차이, 제2 스펙트럼의 발현 차이 및 제1 스펙트럼의 DEP를 검출하기 위한 임계값을 이용하여 제2 스펙트럼의 DEP를 검출하기 위한 임계값을 결정할 수 있다.

발명의 효과

- [0031] 본 발명의 일실시예에 의하면, 동일한 전구체 내에 포함된 펩타이드에 대한 잡음의 정도에 따른 왜곡의 정도를 파악하기 위해 모의 자료를 생성함으로써, 동일한 전구체 내에 존재하는 펩타이드 간의 간섭현상으로 인한 펩타이드의 실제 정량에 따른 잡음의 정도를 보다 명확하게 파악할 수 있다.
- [0032] 본 발명의 일실시예에 의하면, 펩타이드의 실제 정량이 왜곡됨에 따른 비율 확장 또는 비율 축소에 따른 발현차이 펩타이드 검출의 제1종 오류 확률 또는 제2종 오류의 확률을 제어할 수 있는 임계값을 도출함으로써, 전구체 분리 순도에 따른 비율 확장 또는 비율 축소에 대한 위양성(false positive) 혹은 위음성(false negative)으로 추정되는 발현 차이 펩타이드를 최소화할 수 있다.

도면의 간단한 설명

- [0033] 도 1은 본 발명의 일실시예에 따른 펩타이드의 동정 및 정량 분석 방법을 도시한 플로우차트이다.
- 도 2는 본 발명의 일실시예에 따라 4개의 서로 다른 리포터 이온으로 구성된 동중 원소 태그를 이용하여 펩타이드의 정량을 수행하는 과정을 도시한 도면이다.
- 도 3은 본 발명의 일실시예에 따라 샘플로부터 펩타이드에 대한 스펙트럼을 추출하는 과정을 나타낸 도면이다.
- 도 4는 도 3의 예시에 대한 MS 스펙트럼과 MS 스펙트럼에서 추출된 잡음이 포함된 MS/MS 스펙트럼을 나타낸 도면이다.
- 도 5는 본 발명의 일실시예에 따른 원본 스펙트럼을 이용하여 혼합 스펙트럼을 생성하는 과정을 나타낸 도면이다.
- 도 6은 본 발명의 일실시예에 따른 특정 단백질에 대하여 전구체 분리 순도에 따른 펩타이드들의 정량결과를 나타낸 도면이다.
- 도 7은 본 발명의 일실시예에 따른 혼합 스펙트럼에 대한 펩타이드의 동정 결과를 도시한 도면이다.
- 도 8은 본 발명의 일실시예에 따른 비율 확장 오류와 비율 축소 오류를 도시한 도면이다.
- 도 9는 본 발명의 일실시예에 따른 전구체 분리 순도에 따라 발현 차이 펩타이드의 검출을 위한 임계값을 도시한 도면이다.
- 도 10은 본 발명의 일실시예에 따른 임계 값을 기준으로 동정한 발현차이 펩타이드들의 -결과를 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0034] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.
- [0035] 도 1은 본 발명의 일실시예에 따른 펩타이드의 동정 및 정량 분석 방법을 도시한 플로우차트이다.
- [0036] 도 1의 단계(101)는 원본 스펙트럼을 추출하는 과정을 나타낸다. 원본 스펙트럼은 이하의 과정을 통해 추출될 수 있다.
- [0037] 분석 시스템은 질량분석기에 표본들과 표본들을 구분할 수 있는 질량태그가 생성되는 시약을 함께 사용함으로써 표본들간 발현 차이를 확인할 수 있는 스펙트럼들을 생성할 수 있다.
- [0038] 그러면, 분석 시스템은 펩타이드 동정 도구를 이용하여 iTRAQ4plex를 통해 획득된 스펙트럼을 동정할 수 있다. 이 후, 시스템은 전구체 분리 순도(Precursor Isolation Purity:PIP)에 기초하여 동정된 펩타이드들 중 잡음이

포함되지 않은 펩타이드들을 선택할 수 있다. 여기서, 잡음이 포함되지 않은 펩타이드들은 전구체 분리 순도가 100인 것을 의미한다.

[0039] 일례로, 분석 시스템은 FDR(False Discovery Rate)의 1% 내에 속하는 모든 PSM(Peptide-Spectrum-matches)에 대해 전구체 분리 순도를 계산할 수 있다. PSM은 MS/MS 스펙트럼과 해당 스펙트럼에 동정된 펩타이드를 의미한다. 펩타이드의 동정(identification)은 DB내의 존재하는 단백질 서열을 특정 효소로 잘라서 펩타이드들을 만든 후 펩타이드들에 대하여 이론적인 스펙트럼을 생성한다. 분석 시스템은 ITRAQ4plex를 통한 스펙트럼들을 위와 같이 생성된 이론적인 스펙트럼들과 비교함으로써, 해당 스펙트럼에 제일 적합한 펩타이드를 동정한다. FDR은 시스템이 타겟 DB(Target DB)에서 동정한 펩타이드들과 유인 DB(Decoy DB)에서 동정한 펩타이드들 간의 비율을 의미한다.

[0040] 한편, 전구체 분리 순도는 전구체 피크 세기를 전구체 분리 윈도우(Precursor Isolation Window) 내의 전체 피크 세기(total peak intensity)로 나눔으로써 계산될 수 있다. 이 때, 전구체 피크 세기는 전구체 동위원소 클러스터(Precursor Isotope Cluster) 세기의 합으로 결정될 수 있다.

[0041] 데이터베이스 검색 결과로부터 4개의 iTRAQ 리포터 이온 피크들을 가지는 100% 전구체 분리 순도를 가지는 PSM들이 선택될 수 있다. 일례로, 선택된 PSM들은 (3개의) 펩타이드 동정 도구들을 통해 (동일하게) 동정된 펩타이드들을 의미한다. 또한, 선택된 PSM들은 특정 전하량을 나타내며, 미리 설정된 리포터 이온이 모두 존재하여야 한다.

[0042] 이 때, 펩타이드 동정 도구는 MODa, MS-GF+, Proteome Discoverer를 포함한다. 그리고, 선택된 PSM은 사용된 동정 도구에서 모두 동일하게 동정된 펩타이드들이며 +2 또는 +3의 전하량을 나타내고, 4개의 리포터 이온이 모두 존재한다.

[0043] 모의 스펙트럼을 생성하기 위해, 100% 전구체 분리 순도를 가지는 PSM들의 스펙트럼에서 미리 설정된 분리 윈도우(isolation window)에 존재하는 서로 다른 2개의 전구체들(precursors)이 랜덤하게 선택될 수 있다. 여기서, 분리 윈도우는 미리 설정된 전구체의 질량대 전하비[m/z]의 구간에 따라 결정될 수 있다. -

[0044] 단계(105)에서, 분석 시스템은 선택된 2개의 전구체들의 MS/MS 스펙트럼을 미리 설정된 전구체 분리 순도에 따라 혼합함으로써 모의 스펙트럼을 생성할 수 있다. 잡음의 정도가 정량의 왜곡에 미치는 영향을 알아내기 위해서, 분석 시스템은 전구체 분리 순도 지수에 따라 특정 비율을 나타내는 복수의 모의 스펙트럼들을 생성할 수 있다.

[0045] 일례로, 2개의 전구체들의 MS/MS 스펙트럼이 전구체 분리 순도에 따라 혼합되기 전에, 2개의 전구체들의 MS/MS 스펙트럼은 가공될 수 있다.

[0046] 첫째로, 2개의 전구체들의 MS/MS 스펙트럼 각각은 전체 이온 카운트(Total Ion Count: TIC)에 기초하여 정규화될 수 있다. 구체적으로, 2개의 전구체들의 MS/MS 스펙트럼 각각의 TIC가 동일하게 조정된다.

[0047] 둘째로, 2개의 전구체들의 MS/MS 스펙트럼 각각에서 리포터 이온들의 피크 세기가 비닝된다. 일례로, 2개의 전구체들의 MS/MS 스펙트럼 각각에서 각 리포터 이온들의 이론적인 질량을 기준으로 미리 설정된 크기의 윈도우 내에 존재하는 모든 피크를 하나의 피크로 병합함으로써 비닝이 수행된다. 이 때, 미리 설정된 크기의 윈도우는 0.005Da의 윈도우일 수 있다. 이 후, 가공된 2개의 전구체들의 MS/MS 스펙트럼의 모든 피크들이 전구체 분리 순도에 따라 혼합되어 서로 결합된다. 다만, 모든 피크들이 결합되기 전에, 가공된 2개의 전구체들의 MS/MS 스펙트럼을 구성하는 이온 피크의 각 단편(fragment)들이 주어진 전구체 분리 순도에 따라 스케일링될 수 있다. 그리고, 나머지 펩타이드에 대해서는 각 피크의 질량대전하비에 따라 두 개의 정규화된 제1 스펙트럼들의 피크들을 병합할 수 있다.

[0048] 이 때, 가공된 2개의 전구체들의 MS/MS 스펙트럼의 모든 피크들이 혼합되는 경우, 모의 스펙트럼을 의미하는 혼합 스펙트럼(multiplexed spectra)이 생성될 수 있다. 단계(105)는, 분리 시스템이 단계(104)를 통해 생성한 혼합 스펙트럼들을 나타낸다. 이 때, 단계(104)는 단계(101), 단계(102) 및 단계(103) 이후에 진행될 수 있다. 그리고, 단계(105), 단계(106)는 단계(105)가 이후에 진행될 수 있다. 단계(102), 단계(103), 단계(106) 및 단계(107)는 이후에 설명하기로 한다.

[0049] 여기서, 혼합되는 2개의 전구체들의 MS/MS 스펙트럼에서 보다 큰 TIC를 가지는 전구체의 MS/MS 스펙트럼이 참된 펩타이드(true peptide)로부터 도출된 MS/MS 스펙트럼으로 간주되며, 이를 원본 스펙트럼(original spectra)으로 정의할 수 있다.

- [0050] 그리고, 혼합되는 2개의 전구체들의 MS/MS 스펙트럼에서 보다 작은 TIC를 가지는 전구체의 MS/MS 스펙트럼이 거짓 펩타이드(false peptide)로부터 도출된 MS/MS 스펙트럼으로 간주되며, 이를 노이즈 스펙트럼으로 정의할 수 있다. 그러면, 모의 스펙트럼의 전구체에 대한 질량대 전하비는 참된 펩타이드로부터 도출된 스펙트럼의 전구체에 대한 질량대 전하비로 설정될 수 있다.
- [0051] 펩타이드의 동정 및 정량에서 전구체의 세기의 영향을 테스트하기 위해 상기 기준을 만족하는 PSM들이 전구체 세기(low 15%, intermediate 70%, high 15%)에 따라 3가지 빈(bin)으로 분리될 수 있다. 그러면, 혼합 스펙트럼은 같은 빈으로부터 임의의 2개의 스펙트럼을 이용하여 생성될 수 있다.
- [0052] 한편, 본 발명의 일실시예에 의하면, 펩타이드의 동정 및 정량 분석에 있어서 참된 발현 비율(express ratio)의 영향이 확인될 수 있다.
- [0053] 각 PSM에 대해 암-정상 발현 비율(Cancer-to-Normal Expression Ratio)은 iTRAQ 리포터 이온의 세기에 기초하여 계산될 수 있다. 일례로, 4개의 iTRAQ 리포터 이온의 세기가 I114, I115, I116, I117인 경우, 암-정상 발현 비율은 $((I115+I117)/(I114+I116))$ 이 될 수 있다.
- [0054] 위와 같은 암-정상 발현 비율이 계산되기 전에, 동위원소 순도 정정 지수(Isotopic Purity Correction Factor)를 적용함으로써 4개의 iTRAQ 리포터 이온 피크들이 정규화될 수 있다.
- [0055] 시뮬레이션을 위해, 암-정상 발현 비율이 6개의 그룹으로 그룹화될 수 있다. 예를 들어, 6개의 그룹은 [1,2), [2,3), [3,5), [5,10), [10,20), 및 [20,∞) 일 수 있다. 그러면, 각각의 PSM을 암-정상 발현 비율의 그룹에 할당할 때, 암-정상 발현 비율 또는 정상-암 발현 비율이 고려될 수 있다. 특정한 참된 발현 비율 그룹에 대한 모의 스펙트럼은 해당 그룹으로부터 샘플링된 스펙트럼을 임의의 발현 비율을 가지는 노이즈 스펙트럼과 혼합함으로써 생성될 수 있다.
- [0056] < 혼합 스펙트럼으로부터 펩타이드 동정 및 정량>
- [0057] 단계(102)에서, 분석 시스템은 원본 스펙트럼에 대해 펩타이드 동정(peptide identification)을 수행할 수 있다. 유사하게, 단계(106)에서, 분석 시스템은 혼합 스펙트럼에 대해 펩타이드 동정을 수행할 수 있다. 앞서 설명한 MODa, MS-GF+, Proteome Discoverer와 같은 펩타이드 동정 도구가 원본 스펙트럼 및 혼합 스펙트럼에 적용될 수 있다. 펩타이드 동정 이후에, 암-정상 발현 비율이 계산될 수 있다. 이 때, 암-정상 발현 비율은, 동위원소 순도 정정 지수를 적용한 이후에 iTRAQ 리포터 이온 피크들로부터 각각의 혼합 스펙트럼에 대해 계산될 수 있다.
- [0058] 그러면, 단계(103)에서, 분석 시스템은 원본 스펙트럼에 대한 PSM을 추출할 수 있다. 그리고, 단계(107)에서, 분석 시스템은 혼합 스펙트럼에 대한 PSM을 추출할 수 있다. 앞서 살펴본 바와 같이, PSM은 펩타이드 동정에 따라 확인된 펩타이드로서, MS/MS 스펙트럼을 분석하여 데이터베이스를 통해 확인된 펩타이드들을 의미한다. 이 후, 단계(108)에서 분석 시스템은 단계(103) 및 단계(107)을 통해 도출된 PSM을 비교할 수 있다.
- [0059] 단계(109)에서, 분석 시스템은 원본 스펙트럼으로부터 도출된 PSM과 혼합 스펙트럼으로부터 도출된 PSM 간에 일치하는 PSM을 추출할 수 있다. 그리고, 단계(110)에서, 분석 시스템은 원본 스펙트럼에 대해 정량을 수행할 수 있다. 또한, 단계(111)에서, 분석 시스템은 혼합 스펙트럼에 대해 정량을 수행할 수 있다. 이 후, 단계(112)에서, 분석 시스템은 원본 스펙트럼의 정량 결과와 혼합 스펙트럼의 정량 결과를 비교할 수 있다.
- [0060] <Co-Fragmented Spectra로부터 발현 차이 펩타이드의 동정에 대한 임계값을 결정>
- [0061] 혼합 스펙트럼을 시뮬레이션함으로써, 발현 차이 펩타이드(Differentially Expressed Peptides: DEP)의 동정에서 단편화 효과가 정량화될 수 있다. 모의 스펙트럼을 의미하는 혼합 스펙트럼을 이용함으로써, iTRAQ 정량 분석에서 왜곡된 양의 분포가 추정될 수 있다. 여기서, 왜곡된 양은 주어진 전구체 분리 순도에서 의 비율 축소 또는 비율 확장의 정도를 의미한다.
- [0062] 단계(113)에서, 분석 시스템은 원본 스펙트럼의 정량 결과와 혼합 스펙트럼의 정량 결과 간의 차이에 대한 분포를 결정할 수 있다. iTRAQ 정량 분석에서 왜곡된 양은 혼합 스펙트럼과 원본 스펙트럼 간의 암-정상 발현 비율의 차이로 정의될 수 있다. iTRAQ 정량 분석에서 왜곡된 양의 분포를 추정하기 위해, 원본 스펙트럼의 펩타이드와 동일한 펩타이드에 매칭되는 혼합 스펙트럼이 이용될 수 있다.
- [0063] 단계(114)에서, 분석 시스템은 Co-fragmented 스펙트럼에 대해 발현 차이 펩타이드를 검출하기 위한 임계값을 단계(113)에서 도출된 분포를 이용하여 계산 후 Co-Fragmented 펩타이드들에 적용하여 발현 차이 펩타이드를 결

정할 수 있다. 여기서, Co-Fragmented 펩타이드는 전구체 분리 순도가 100미만인 경우를 의미한다. 구체적으로, 주어진 전구체 분리 순도에 대해 스펙트럼이 iTRAQ 정량 분석에서 보이는 왜곡된 양의 분포에 기반하여 발현 차이 펩타이드의 동정을 위한 임계값(Cut-off value)이 계산될 수 있다.

[0064] 이러한 임계값은 DEP들 중 위양성(false positive) DEP를 제어하기 위해 사용될 수 있다. DEP 동정에서 제1종 오류(위양성)는 혼합 스펙트럼에서 동정되는 DEP가 원본 스펙트럼에서는 동정되지 않는 경우를 의미한다.

[0065] 주어진 제1종 오류 확률에서, DEP를 위한 임계값은 다음과 같이 결정될 수 있다.

[0066] 각 스펙트럼에 대해 암-정상 발현 비율이 계산될 수 있다. 그러면, 각 암-정상 발현 비율에 대해 로그가 적용될 수 있다. MR은 100%보다 낮은 전구체 분리 순도를 가지는 혼합 스펙트럼에 대해 로그로 표현된 암-정상 발현 비율을 의미한다. OR은 100% 전구체 분리 순도를 가지는 원본 스펙트럼에 대해 로그로 표현된 암-정상 발현 비율을 의미한다.

[0067] 한편, DR은 OR과 MR사이의 차이를 의미한다. OC1(lower)과 OC2(upper)는 원본 스펙트럼으로부터 DEP 동정을 위한 임계값을 의미한다. 일례로, 2-fold DEP 동정에서, OC1는 -1이고, OC2는 1일 수 있다. 그리고, MC1(lower), MC2(upper)는 특정 전구체 분리 순도를 가지는 혼합 스펙트럼으로부터 DEP 동정을 위한 임계값을 의미한다. 그러면, MC1과 MC2를 이용하여 혼합 스펙트럼으로부터 DEP 동정을 수행할 때, 제1종 오류의 확률은 하기 수학적 식 1에 따라 계산될 수 있다.

수학적 식 1

$$\begin{aligned} & P(O_R \geq OC_1, M_R < MC_1) + P(O_R \leq OC_2, M_R > MC_2) \\ & = P(O_R - M_R > OC_1 - MC_1) \\ & + P(O_R - M_R < OC_2 - MC_2) \\ & = P(D_R > OC_1 - MC_1) + P(D_R < OC_2 - MC_2) \end{aligned}$$

[0068]

[0069] 상기 수학적 식 1에 의해, 주어진 제1종 오류의 확률(ex, p-value: 0.05)를 이용하여 MC1과 MC2가 결정될 수 있다. 그리고, 결정된 MC1과 MC2는 단편화된 스펙트럼으로부터 강건한 DEP 검출을 위해 이용될 수 있다.

[0070] 유사한 방식으로, 위음성(false negative) DEP를 제어하기 위한 임계값도 도출될 수 있다. DEP 동정에서 제2종 오류(위음성)는 혼합 스펙트럼에서 동정되지 않는 DEP가 원본 스펙트럼에서는 동정되는 경우를 의미한다. 그러면, MC1과 MC2를 이용하여 혼합 스펙트럼으로부터 DEP 동정을 수행할 때, 제2종 오류의 확률은 하기 수학적 식 2에 따라 계산될 수 있다.

수학적 식 2

$$\begin{aligned} & P(O_R < OC_1, M_R \geq MC_1) + P(O_R > OC_2, M_R \leq MC_2) \\ & = P(O_R - M_R < OC_1 - MC_1) + P(O_R - M_R > OC_2 - MC_2) \\ & = P(D_R < OC_1 - MC_1) + P(D_R > OC_2 - MC_2) \end{aligned}$$

[0071]

[0072] 2-fold DEP 동정의 성능을 확인하기 위해, 혼합 스펙트럼이 사용될 수 있다. 2-fold DEP 동정은 아래 방법에 의해 수행된다.

[0073] (i) S2I(Signal-to-Interference)를 이용한 비율 정정(ratio correction): RATIO_COR

[0074] (ii) 종래의 임계값: -1 및 1

[0075] (iii) 제1종 오류의 확률을 정정하기 위한 임계값: CUTOFF_COR

[0076] DEP 동정의 성능은 긍정 예측값(Positive Predictive Value: PPV), 민감도 및 F1값에 의해 평가될 수 있다.

여기서, PPV는 (# of True Positive) / (# of True Positive + # of False Positive)로 정의될 수 있다. 즉, PPV는 동정된 DEP에서 참된 DEP의 비율을 의미한다. 민감도는 (# of True Positive) / (# of True Positive + # of False Negative)로 정의될 수 있다. F1값은 PPV와 민감도의 조화 평균을 의미한다.

[0077] 다만, 혼합 스펙트럼에 대해서 크로마토그램 정보가 이용 가능하지 않으므로, S2I가 계산될 수 없다. 그래서, RATIO_COR 대신에 전구체 분리 순도가 이용될 수 있다.

[0078] 앞서 설명한 도 1의 과정을 요약하면 다음과 같다.

[0079] 분석 시스템은 잡음이 없는 원본 스펙트럼을 이용하여 혼합 스펙트럼을 생성한다. 여기서, 잡음이 없는 스펙트럼은 4개의 리포터 이온들을 모두 포함한다. 혼합 스펙트럼은 시물레이팅될 수 있으며, 모의 스펙트럼으로 정의될 수 있다.

[0080] 혼합 스펙트럼을 생성하기 위해 이용되는 원본 스펙트럼 및 혼합 스펙트럼은 복수의 펩타이드 동정 도구를 이용하여 동정된다. 펩타이드 동정이 완료된 이후에, 부정확하게 동정된 혼합 스펙트럼을 필터링한다. 그리고, 원본 스펙트럼과 혼합 스펙트럼 각각의 암-정상 발현 비율의 로그값이 비교되며, 이에 따라 암-정상 발현 비율의 로그값 차이의 분포가 결정된다. 그러면, 주어진 전구체 분리 순도와 제1종 오류 확률에 대해 혼합 스펙트럼을 위한 DEP 동정의 임계값이 결정될 수 있다.

[0081] 결론적으로 도 1에 의하면, 동중 원소 태그를 사용한 정량 자료 중 잡음이 포함되지 않은 원본 스펙트럼들을 혼합하여 인위적으로 잡음이 포함된 혼합 스펙트럼들이 생성된다. 그러면, 분석 시스템은 생성된 혼합 스펙트럼을 이용하여 잡음의 정도가 펩타이드의 정량의 왜곡에 미치는 영향을 확인할 수 있다. 이러한 잡음은 함께 단편화됨으로써 발생되며, 잡음으로 인한 영향은 확률 분포로 표현될 수 있다. 그러면, 분석 시스템은 확률 분포를 이용하여 비율 축소나 비율 확장의 확률을 통제할 수 있는 발현 차이 펩타이드(differentially-expressed peptide, DEP)를 추출하기 위한 임계값을 계산할 수 있다.

[0082] 도 2는 본 발명의 일실시예에 따라 4개의 서로 다른 리포터 이온으로 구성된 동중 원소 태그를 이용하여 펩타이드의 정량을 수행하는 과정을 도시한 도면이다.

[0083] 도 2는 iTRAQ 정량 분석 과정을 도식화한 것이다. 여기서, 동중 원소 태그는 리포터 이온과 밸런스 태그로 구성될 수 있다. 리포터 이온의 질량은 각각 114, 115, 116, 117을 나타내며, 밸런스 태그의 질량은 각각 31, 30, 29, 28을 나타낸다. 이러한 동중 원소 태그는 서로 다른 조건에서 추출한 샘플들 각각에 적용된다. 실제로 4개의 샘플들 각각에 적용되는 리포터 이온의 질량은 서로 다르지만, 서로 다른 리포터 이온들과 밸런스 태그와 결합된 동중 원소 태그는 동일한 질량을 나타낸다. 그러면, 동중 원소 태그가 결합된 4개의 샘플들을 결합하여 MS(Mass Spectrometry)를 적용하면 하나의 피크가 생성된다. 이러한 피크가 전구체를 나타낸다.

[0084] 이 후, MS/MS를 적용하면 동중 원소 태그에 결합된 밸런스 태그가 제거된다. 하면, 그러면, 샘플들 각각에 결합된 질량이 다른 리포터 이온들이 추출된다. 하지만, 샘플들로부터 추출되는 펩타이드의 단편(fragment)들은 동일한 결과를 나타낸다.

[0085] 즉, iTRAQ를 통한 정량 분석은 펩타이드의 말단에 동일한 질량대 전하비를 나타내는 동중 원소 태그를 서로 다른 조건에서 추출한 샘플들 각각에 라벨링하는 분석 기법을 의미한다. 동중 원소 태그에 포함된 리포터 그룹과 밸런스 그룹의 질량의 합이 동일하기 때문에, MS 스펙트럼에서 4개의 샘플이 하나의 피크로 검출된다.

[0086] 이 후, MS/MS 스펙트럼에서, 리포터 그룹에 해당하는 피크들이 리포터 이온의 질량에 따라 114 내지 117 사이의 질량대 전하비 범위에서 나타난다. 그리고, 리포터 이온에 대응하는 피크들 각각의 세기(intensity)는 리포터 이온에 결합된 샘플들의 펩타이드의 양을 나타낸다. 앞서 살펴본 바와 같이, 샘플들은 서로 다른 조건에서 추출된다.

[0087] 결과적으로, iTRAQ 정량 분석은 동중 원소 태그에 포함된 리포터 이온들 각각의 피크를 이용하여 피크 영역을 비교하여 피크 영역을 비율을 결정함으로써 서로 다른 조건에서 발현된 펩타이드들 간의 상대적인 정량 결과를 도출할 수 있다. 즉, iTRAQ 정량 분석을 통해 서로 다른 조건에 있는 단백질에 해당하는 펩타이드의 상대적인 양으로, 펩타이드의 발현 정도가 측정될 수 있다. 그리고, 4개의 샘플들로부터 추출된 펩타이드들은 서로 같은 fragment 패턴을 가지므로, MS/MS 스펙트럼을 통해 펩타이드의 아미노산 서열을 확인함으로써 정량 결과와 함께 펩타이드의 동정도 동시에 수행될 수 있다.

[0088] 도 3은 본 발명의 일실시예에 따라 샘플로부터 펩타이드에 대한 스펙트럼을 추출하는 과정을 나타낸 도면이다.

- [0089] 도 3은 복수의 샘플들이 혼합된 결과로부터 MS와 MS/MS를 수행한 결과를 나타낸다. 복수의 샘플들이 혼합된 결과에서 MS를 수행하면, 각 샘플들로부터 추출된 펩타이드에 대응하는 피크들이 MS 스펙트럼에 표시될 수 있다. 여기서, 가는 실선이 관심있는 전구체(interest precursor)이고, 굵은 실선이 노이즈라고 가정한다.
- [0090] 그러면, MS/MS가 수행되면 MS/MS 스펙트럼에 관심있는 전구체로부터 도출된 결과에 노이즈가 결합되어 결과가 왜곡될 수 있다.
- [0091] 도 4는 도 3의 예시에 대한 MS 스펙트럼과 잡음이 포함된 MS/MS 스펙트럼을 나타낸 도면이다.
- [0092] 도 4에서 MS1은 MS 스펙트럼을 나타낸다. MS1에서 볼 수 있듯이, 관심있는 전구체(굵은 실선)와 노이즈(가는 실선)가 서로 유사한 질량대 전하비를 가져서 MS 스펙트럼의 분리 윈도우 범위(점선) 내에 속할 수 있다. 그러면, MS/MS를 통해 전구체와 노이즈가 함께 단편화됨(co-fragmented)으로써 관심있는 전구체의 결과에 노이즈가 포함될 수 있다.
- [0093] MS2에서 볼 수 있듯이, MS/MS 스펙트럼에서 관심있는 전구체(굵은 실선)와 노이즈(가는 실선)가 서로 분리되지 않고 조합된 결과가 도출될 수 있다. 그러면, 관심있는 전구체에 대한 왜곡된 리포터 이온의 피크가 생성되며, 이로 인해서 펩타이드 동정 및 정량의 결과에 왜곡이 발생할 수 있다.
- [0094] 도 5는 본 발명의 일실시예에 따른 원본 스펙트럼을 이용하여 혼합 스펙트럼을 생성하는 과정을 나타낸 도면이다.
- [0095] 도 5에서 제1 스펙트럼은 잡음이 없는 원본 스펙트럼(스펙트럼 A, 스펙트럼 B)을 나타낸다. 분석 시스템은 선택된 2개의 전구체들의 MS/MS 스펙트럼을 미리 설정된 전구체 분리 순도에 따라 혼합함으로써 모의 스펙트럼인 제2 스펙트럼을 생성할 수 있다. 일례로, 2개의 전구체들의 MS/MS 스펙트럼이 전구체 분리 순도에 따라 혼합되기 전에, 2개의 전구체들의 MS/MS 스펙트럼은 가공될 수 있다.
- [0096] 먼저, 2개의 전구체들의 MS/MS 스펙트럼 각각은 전체 이온 카운트(Total Ion Count: TIC)에 기초하여 정규화될 수 있다. 구체적으로, 2개의 전구체들의 MS/MS 스펙트럼 각각의 TIC가 동일하게 조정된다.
- [0097] 이후, 2개의 전구체들의 MS/MS 스펙트럼 각각에서 리포터 이온들의 피크 세기가 비닝된다. 일례로, 2개의 전구체들의 MS/MS 스펙트럼 각각에서 미리 설정된 크기의 윈도우마다 존재하는 2개의 리포터 이온들의 피크를 하나의 피크로 병합함으로써 비닝이 수행된다. 이 때, 미리 설정된 크기의 윈도우는 0.005Da의 윈도우일 수 있다.
- [0098] 이 후, 가공된 2개의 전구체들의 MS/MS 스펙트럼의 모든 피크들이 전구체 분리 순도에 따라 혼합되어 서로 결합된다. 다만, 모든 피크들이 결합되기 전에, 가공된 2개의 전구체들의 MS/MS 스펙트럼을 구성하는 이온 피크의 각 단편(fragment)들이 주어진 전구체 분리 순도에 따라 스케일링될 수 있다.
- [0099] 이 때, 가공된 2개의 전구체들의 MS/MS 스펙트럼의 모든 피크들이 혼합되는 경우, 모의 스펙트럼을 의미하는 혼합 스펙트럼(multiplexed spectra)이 생성될 수 있다.
- [0100] 도 6은 본 발명의 일실시예에 따른 특정 단백질에 대하여 전구체 분리 순도에 따른 펩타이드들의 정량결과를 나타낸 도면이다. 도 6을 참고하면, 비율 확장의 결과로부터, 수정된 임계값(Cutoff value)이 위양성 DEP들을 동정하고 감소시킬 수 있다는 것을 확인할 수 있다. 구체적으로, 7개의 고유한 시퀀스 펩타이드들 중 5개 시퀀스 펩타이드들은 전구체 분리 순도(PIP)가 100%인 MS/MS 스펙트럼을 가진다. 이 때, 5개의 시퀀스 펩타이드들의 리포터 이온의 스펙트럼은 평균 1.19의 fold change를 나타낸다.
- [0101] 도 6(a)는 단백질 Q9BX68의 아미노산 시퀀스를 나타낸다. 구체적으로, 도 6(a)의 경우 단백질 Q9BX68에 대해 동정된 5개의 고유한 시퀀스 펩타이드들로 동정한 결과를 나타낸다. 이에 대응하는 iTRAQ 리포터 이온 스펙트럼이 관측된 fold change와 함께 도시된다.
- [0102] 도 6(b)은 분리 윈도우 내에서 전구체 이온의 MS 스펙트럼을 확대한 것이다. 타겟 전구체 이온은 검은 원으로 표시된 것으로, 회색 원으로 표시된 다른 이온과 함께 단편화된다. 타겟 전구체 이온의 전구체 분리 순도(PIP)는 20%이다.
- [0103] 전구체 분리 순도가 20%일 때, 시퀀스 펩타이드인 SLPADILYEDQQCLVFR(검은원)의 리포터 이온의 스펙트럼은 평균 2.32의 fold change를 나타낸다. 비록, fold change가 2보다 더 큰 것처럼 보이지만, 전구체 분리 순도가 50%일 때 DEP에 대한 수정된 임계값은 3.94이다. 그래서, 펩타이드들이 DEP로 동정되지 않고, 위양성(false positive)으로 추정되는 DEP로서 제거된다.

- [0104] SLPADILYEDQQCLVFR는 단백질 Q9BX68의 펩타이드 시퀀스를 의미한다. 단백질 Q9BX68은 동형 단백질(isoform)이 아니기 때문에, 비율 확장은 동형 단백질 혼합(isoform mixing)에 의해 발생되지 않을 것으로 판단된다. 동형 단백질 혼합은 비율 변환(ratio change)에서 2 이상의 동형 단백질이 다르게 발현된 결과를 의미한다. 비율 확장은 VFIPVLQSVTA(회색원) 펩타이드의 함께 단편화된 결과로 인해 야기될 수 있다. VFIPVLQSVTA는 전구체 분리 순도가 84%에서 리포터 이온의 스펙트럼은 2.73의 fold change를 나타낸다.
- [0105] 도 6(c)는 iTRAQ 리포터 이온 스펙트럼을 가지는 함께 단편화된 2개의 이온의 MS/MS 스펙트럼이 도시된다. 계산된 fold change는 2.32이다. 함께 단편화된 스펙트럼(cofragmented spectrum)은 2개의 단편화된 펩타이드들(SLPADILYEDQQCLVFR 및 VFIPVLQSVTAR)로부터 도출된 단편에 의해 설명된다.
- [0106] 도 7은 본 발명의 일실시예에 따른 혼합 스펙트럼에 대한 펩타이드의 동정 결과를 도시한 도면이다.
- [0107] 도 7은 혼합 스펙트럼으로부터 펩타이드 동정 도구(MoDa 및 MS-GF+)를 이용하여 펩타이드 동정된 결과를 나타낸다. 도 7에 의하면, 혼합 스펙트럼을 이용하여 펩타이드 동정에서 함께 단편화(cofragmentation)의 효과를 알 수 있다. x축은 전구체 분리 순도를 의미하며, y축은 동정된 PSM의 개수를 나타낸다.
- [0108] 혼합 스펙트럼에서 동정된 펩타이드와 원본 스펙트럼에서 동정된 펩타이드가 같은 경우, Correct ID가 설정된다. 그리고, 혼합 스펙트럼에서 동정된 펩타이드와 원본 스펙트럼에서 동정된 펩타이드가 다른 경우, Incorrect ID(Noise)가 결정된다. 다른 잘못된 동정은 Incorrect ID(Other)가 결정된다. 괄호는 동정된 전체 PSM에 대한 비율을 의미한다.
- [0109] 도 7은 인위적으로 혼합한 제2 스펙트럼들을 사용하여 펩타이드 동정 도구인 MoDa 및 MS-GF+에 대해, 전구체 분리 순도에 따라 FDR이 1% 범위에서 동정된 PSM의 개수가 어떻게 변화하는 지를 나타낸다. 전구체 분리 순도가 증가할수록 동정된 PSM의 개수는 증가한다. MoDa에 대해 전구체 분리 순도가 70% 이상인 경우, 동정된 PSM의 비율은 99.1%(7429/7500)가 된다. 그리고, MS-GF+에 대해 전구체 분리 순도가 50% 이상인 경우 동정된 PSM의 비율은 99.9%(7494/7500)이 된다.
- [0110] 도 7에서 설명되는 동정 결과는 정확도로 평가될 수 있다. 만약, 원본 스펙트럼에 대한 PSM이 혼합 스펙트럼에 대한 PSM과 일치하는 경우, 혼합 스펙트럼에 대한 PSM은 정확한 것으로 고려된다. 전구체 분리 순도가 80% 이상인 경우, MoDa에 대해 99% 이상으로 PSM이 정확하게 동정될 수 있다.
- [0111] 그리고, 전구체 분리 순도가 70% 이상인 경우, MG-GF+에 대해 99% 이상으로 PSM이 정확하게 동정될 수 있다. MoDa와 MG-GF+에 대해 부정확하게 동정된 결과의 대부분은 노이즈 스펙트럼과 매칭된다. 따라서, 전구체 분리 순도가 80% 이상인 혼합 스펙트럼으로부터 펩타이드 동정이 수행되면, 적어도 1% 미만의 오류율을 나타낼 것이다.
- [0112] 도 8은 본 발명의 일실시예에 따른 비율 확장 오류와 비율 축소 오류를 도시한 도면이다.
- [0113] 도 1에서 살펴본 바와 같이, iTRAQ 기반의 정량 분석에 있어서 함께 단편화된 결과를 조사하기 위해서, 혼합 스펙트럼으로부터 정확하게 동정된 PSM이 사용된다. 여기서, 혼합 스펙트럼으로부터 정확하게 동정된 PSM은 혼합 스펙트럼으로부터 동정된 PSM들 중 원본 스펙트럼으로부터 동정된 PSM과 일치하는 PSM을 의미한다.
- [0114] 만약, 원본 스펙트럼이 혼합되기 전에, 혼합 스펙트럼의 비율이 원본 스펙트럼의 비율보다 1에 보다 가까운 경우, 혼합 스펙트럼은 비율 축소(ratio compressed)로 정의될 수 있다.

표 1

Ratio compression	Precursor isolation purity							Mean
	50%	60%	70%	80%	90%	95%	99%	
Yes	4765	5246	5416	5484	5510	5518	5520	5351
	(73.50%)	(73.58%)	(73.64%)	(73.62%)	(73.59%)	(73.60%)	(73.61%)	(73.59%)
No	1718	1884	1939	1965	1977	1979	1979	1920
	(26.50%)	(26.42%)	(26.36%)	(26.38%)	(26.41%)	(26.40%)	(26.39%)	(26.41%)
Total	6483	7130	7355	7449	7487	7497	7499	7271
	(100.00%)	(100.00%)	(100.00%)	(100.00%)	(100.00%)	(100.00%)	(100.00%)	(100.00%)

[0115]

- [0116] 표 1은 MoDa에 의해 정확하게 동정된 혼합 스펙트럼을 나타낸다. 표 1은 다양한 전구체 분리 순도에서 비율 축

소 스펙트럼의 비율과 개수를 나타낸다. 표 1에 의하면, MODa에 의해 정확하게 동정된 혼합 스펙트럼의 약 74%가 비율 축소인 것을 알 수 있다. 반대로, 정확하게 동정된 혼합 스펙트럼의 25% 이상이 비율 확장이라는 것을 알 수 있다.

- [0117] 혼합 스펙트럼을 이용함으로써 비율 축소와 비율 확장이 정량화될 수 있다. 이와 관련하여, 도 8은 노이즈 레벨에 따라 비율 축소와 비율 확장이 어떻게 나타나는지를 도시한다. 여기서, 노이즈 레벨은 전구체 분리 순도에 대응한다. 그리고, 비율 축소와 비율 확장의 정도는 혼합 스펙트럼과 원본 스펙트럼 간에 로그 2 fold-change ratio 차이 분포(DR)에 대응한다.
- [0118] 도 8에서 A, C는 MODa에 관한 것이고, B, D는 MS-GF+에 관한 것이다. A, B는 비율 축소를 나타내는 혼합 스펙트럼을 도시한 것이고, C, D는 비율 확장을 나타내는 혼합 스펙트럼을 도시한 것이다. 도 8의 그래프는 혼합 스펙트럼의 전구체 분리 순도에 따라 도시된다.
- [0119] 비율 축소의 DR 평균은 MODa에 대해 0.068에서 0.001 사이의 값을 나타내고, MS-GF+에 대해 0.071에서 0.001의 값을 나타낸다. 그리고, 비율 확장의 DR 평균은 MODa에 대해 -0.103에서 -0.003 사이의 값을 나타내고, MS-GF+에 대해 -0.009에서 -0.003의 값을 나타낸다. 이들 평균값은 노이즈 레벨인 전구체 분리 순도가 감소함에 따라 0에 가까운 값을 나타낸다. 이것은 함께 단편화된 효과를 나타내는 DR 평균이 전구체 분리 순도에 반비례하는 것을 의미한다. 다만, 전구체 분리 순도가 증가할수록 DR의 표준편차는 감소하는 것을 나타낸다.
- [0120] 예를 들어, 전구체 분리 순도가 50%인 경우, 혼합 스펙트럼에 대해 DR의 표준편차는 비율 축소일 때 0.8, 비율 확장일 때 0.4를 나타낸다. 그리고, 전구체 분리 순도가 99%인 경우, 혼합 스펙트럼에 대해 DR의 표준편차는 비율 축소일 때 0.038, 비율 확장일 때 0.013를 나타낸다.
- [0121] 비율 축소에 대한 DR의 분포는 비율 확장일 때보다 더 넓게 나타난다. 이러한 결과는, 함께 단편화된 효과가 비율 확장일 때보다 비율 축소일 때 일반적으로 더 크기 때문이다.
- [0122] 도 9는 본 발명의 일실시예에 따른 전구체 분리 순도에 따라 발현 차이 펩타이드의 검출을 위한 임계값을 도시한 도면이다.
- [0123] 도 9에서, MC1은 DEP 검출을 위한 하위 임계값(lower cutoff value)을 의미하고, MC2는 DEP 검출을 위한 상위 임계값(upper cutoff value)을 의미한다. 도 9를 참고하면, MODa와 MS-GF+에 대해 정정된 임계값이 도시된다. 전구체 분리 순도가 99%일 때, MODa 및 MS-GF+에 대한 MC1과 MC2는 각각 -1.02와 1.03을 나타낸다. 도 9에 의하면, 전구체 분리 순도가 낮아질수록 임계값의 절대값은 증가한다.
- [0124] 전구체 분리 순도가 50%인 혼합 스펙트럼에 대한 MODa의 임계값은 -1.91과 2가 되고, MS-GF+의 임계값은 -1.93과 1.98이 된다. 이는, 같은 양을 나타내는 서로 다른 전구체들이 함께 단편화되는 경우, 4-fold 발현 차이 이상을 나타내는 PSM만이 실제로 2-fold 발현 차이로 고려된다는 것을 의미한다.
- [0125] 도 10은 본 발명의 일실시예에 따른 서로 다른 동정 도구에서 발현 차이 펩타이드의 동정 결과를 도시한 도면이다.
- [0126] 도 10(a)는 제1종 오류 확률이 0.05인 경우, 정정된 임계값을 이용한 (A) MODa에 대한 2-fold DEP 동정 결과를 나타낸다. 그리고, 도 10(b)는 제1종 오류 확률이 0.05인 경우, 정정된 임계값을 이용한 MS-GF+에 대한 2-fold DEP 동정 결과를 나타낸다. x축은 스펙트럼에서 전구체 분리 순도의 범위를 나타낸다. 그리고, y축은 동정된 DEP 개수를 의미한다. 여기서, 추정 위양성(putative false positives)은 일반적인 임계값 -1과 1에 의해 동정된 DEP 개수를 의미한다. 괄호 속의 비율은 동정된 전체 펩타이드 개수와 비율을 의미한다.
- [0127] 본 발명의 일실시예에 의하면, 같은 샘플로부터 도출된 함께 단편화된 혼합 스펙트럼에 다양한 전구체 분리 순도에 대한 정정된 임계값을 적용할 수 있다. 전구체 분리 순도가 50%이상인 경우, 스펙트럼으로부터 도출된 PSM의 개수는 MODa에 대해 381,603을 나타내고, MS-GF+에 대해 439,808을 나타낸다.
- [0128] 도 10은 MODa 및 MS-GF+에 대해 제1종 오류 확률이 0.05일 때 전구체 분리 순도에 따라 정정된 임계값과 일반적인 log2 비율의 임계값(-1, 및 1)에 의해 동정된 2-fold DEP의 개수 변화를 나타낸다. 정정된 임계값이 스펙트럼에 적용될 때, 실제 전구체 분리 순도보다 작거나 같은 전구체 분리 순도에 대한 정정된 임계값이 사용된다. 예를 들어, 실제 전구체 분리 순도가 80%일 때의 정정된 임계값은 실제 전구체 분리 순도가 85%인 스펙트럼에 적용될 수 있다.
- [0129] 함께 단편화되지 않아서 전구체 분리 순도가 100%인 스펙트럼으로부터 동정된 DEP의 개수는 MODa에 대해 9891이

고, MS-GF+에 대해 11,234를 나타낸다. 그리고, Proteome Discoverer에 대해 동정된 DEP의 개수는 8,257을 나타낸다. 이 때, 스펙트럼으로부터 동정된 DEP에는 위양성으로 추정된 DEP가 포함되지 않는다.

[0130] 여기서, 정정된 임계값이 아닌 일반적인 임계값을 적용함으로써 동정된 DEP들을 위양성으로 추정된 DEP라고 정의한다. 스펙트럼에서 노이즈 레벨이 증가할수록, 일반적인 임계값에 의해 동정된 DEP의 개수 및 위양성으로 추정된 DEP의 개수도 증가한다.

[0131] 그러나, 전구체 분리 순도가 70% 미만인 경우, 정정된 임계값에 의해 탐지된 DEP의 개수는 증가되는 정도가 침체된다. MODa에 대해 전구체 분리 순도가 70% 이상인 스펙트럼으로부터 동정된 DEP의 개수는 14,975이다. 그러나, 전구체 분리 순도가 50% 미만인 스펙트럼으로부터 동정된 DEP의 개수는 단지 278을 나타낸다. MS-GF+와 Proteome Discoverer의 결과도 MODa일 때와 유사하다.

[0132] 결론적으로, 전구체 분리 순도가 70% 미만인 스펙트럼은 너무 노이즈가 심해서 2-fold DEP을 검출하기 어려울 수 있다. 하지만, 본 발명의 일실시예에 따르면, 제1종 오류와 노이즈 레벨에 따라 정정된 임계값을 적용함으로써, 노이즈가 포함된 스펙트럼이라고 하더라도 진양성(true positive)일 것으로 판단되는 DEP를 동정할 수 있다.

[0133] 위에서 언급한 과정들을 정리하면 다음과 같다.

[0134] iTRAQ와 같이 동중 원소 태그 기반의 정량 방법은 MS/MS 단계에서 함께 단편화(cofragmentation)됨으로써 펩타이드의 정량 및 동정 결과의 오류가 발생할 수 있다. 이와 같이, 함께 단편화된 효과를 확인하기 위해, 본 발명의 일실시예에 따르면 분석 시스템은 미리 알려진 전구체 구성과 다양한 전구체 분리 순도를 가지는 혼합 스펙트럼을 시뮬레이션할 수 있다.

[0135] 펩타이드 동정을 위해 일반적으로 사용되는 펩타이드 동정 도구가 전구체 분리 순도가 80% 이상을 나타내는 혼합 스펙트럼에 적용될 수 있다. iTRAQ 정량 결과로부터 비율 축소가 확인될 수 있다. 또한, 혼합 스펙트럼에 대한 iTRAQ 정량 결과로부터 비율 확장도 확인될 수 있다.

[0136] 본 발명의 일실시예에 따르면, iTRAQ 정량에서 함께 단편화된 효과를 추정하기 위해, 다양한 전구체 분리 순도와 제1종 오류 비율에서의 정정된 임계값이 계산될 수 있다. 정정된 임계값을 혼합 스펙트럼에 적용함으로써, 분석 시스템은 비율 확장으로 인해 발생할 수 있는 위양성으로 추정된 DEP들을 효과적으로 제거할 수 있다. 그리고, 노이즈가 포함된 스펙트럼을 노이즈 레벨에 따라 필터링함으로써, 적절한 개수의 진양성 DEP가 도출될 수 있다. 본 발명의 일실시예에 따르면, 혼합 스펙트럼을 이용함으로써 iTRAQ의 정량 분석 과정에서 함께 단편화로 인한 효과를 용이하게 확인할 수 있다.

[0137] 전구체 분리 순도가 커짐에 따라 PSM을 동정한 결과의 정확도는 증가할 수 있다. 그리고, 전구체 분리 순도가 80%이상이면 스펙트럼의 동정 결과는 1% 미만의 오류만 존재한다. 그리고, 함께 단편화된 스펙트럼의 정량 결과가 실제 정량 결과보다 더 큰 “비율 확장” 현상을 보이는 스펙트럼은 26%를 나타낸다.

[0138] 또한, “비율 확장”된 스펙트럼들을 대상으로 정정된 임계값을 위암 데이터에 적용하여 DEP를 추출한 결과, 2배 차이나는 펩타이드(2-fold DEP)의 검출 결과는 전구체 분리 순도가 증가함에 따라 오류가 줄어드는 것을 알 수 있다. 그리고, 전구체 분리 순도가 70%이상이면, 위양성으로 추정되는 DEP가 25% 정도 감소되는 것을 알 수 있다.

[0139] 본 발명의 실시 예에 따른 방법들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0140] 이상과 같이 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상의 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.

[0141] 그러므로, 본 발명의 범위는 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

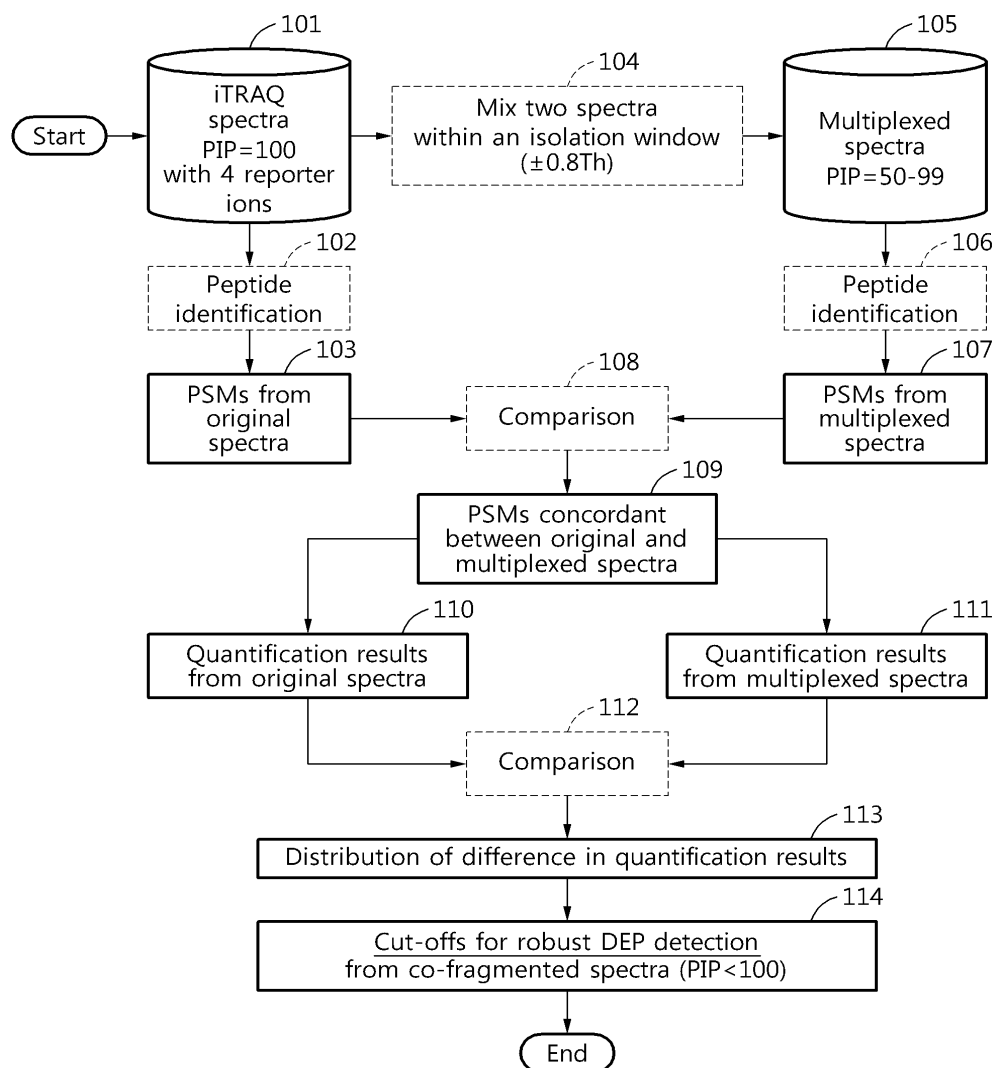
부호의 설명

[0142]

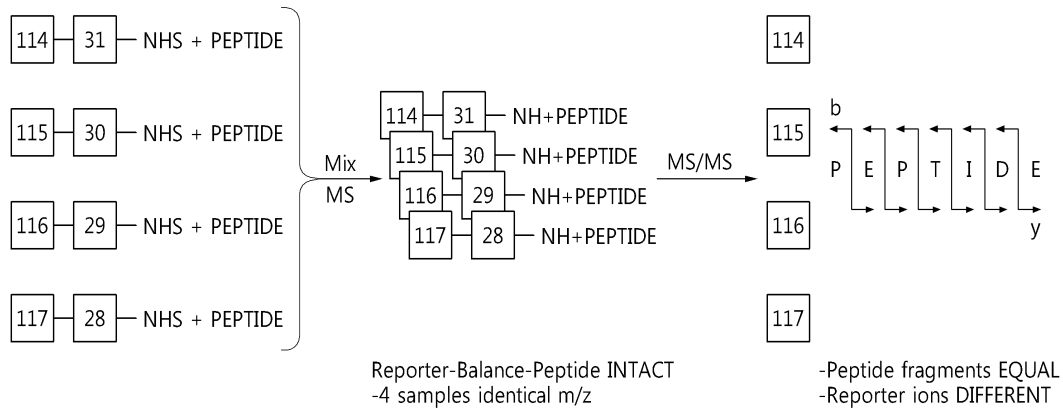
- 101: 원본 스펙트럼
- 102: 원본 스펙트럼에 대한 펩타이드 동정
- 103: 원본 스펙트럼으로부터 PSM 추출
- 104: 원본 스펙트럼의 혼합
- 105: 모의 스펙트럼
- 106: 모의 스펙트럼에 대한 펩타이드 동정
- 107: 모의 스펙트럼으로부터 PSM 추출

도면

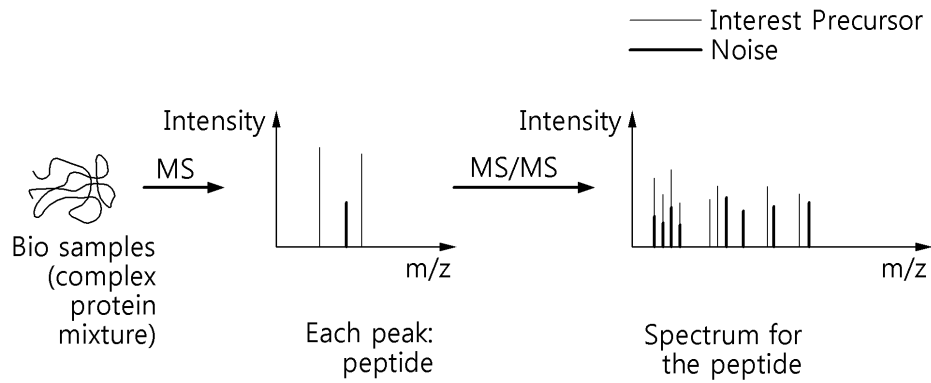
도면1



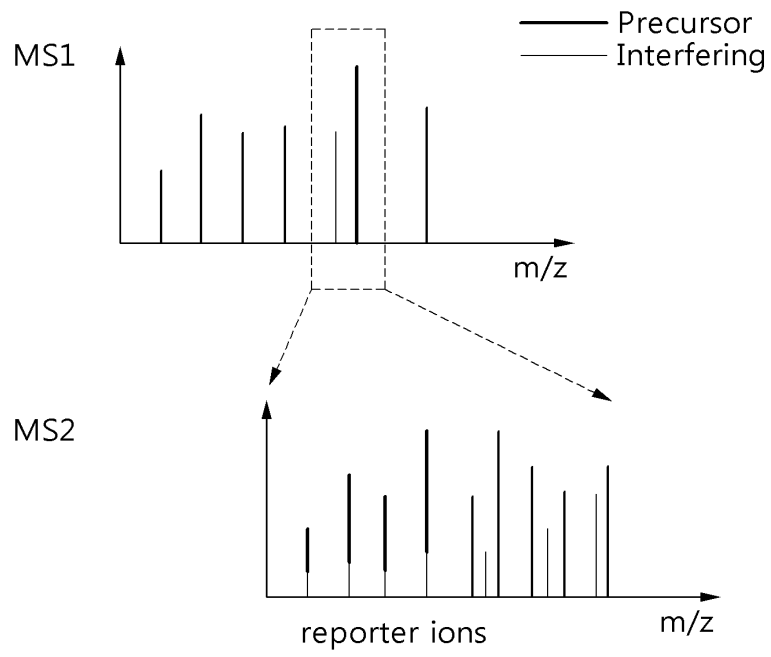
도면2



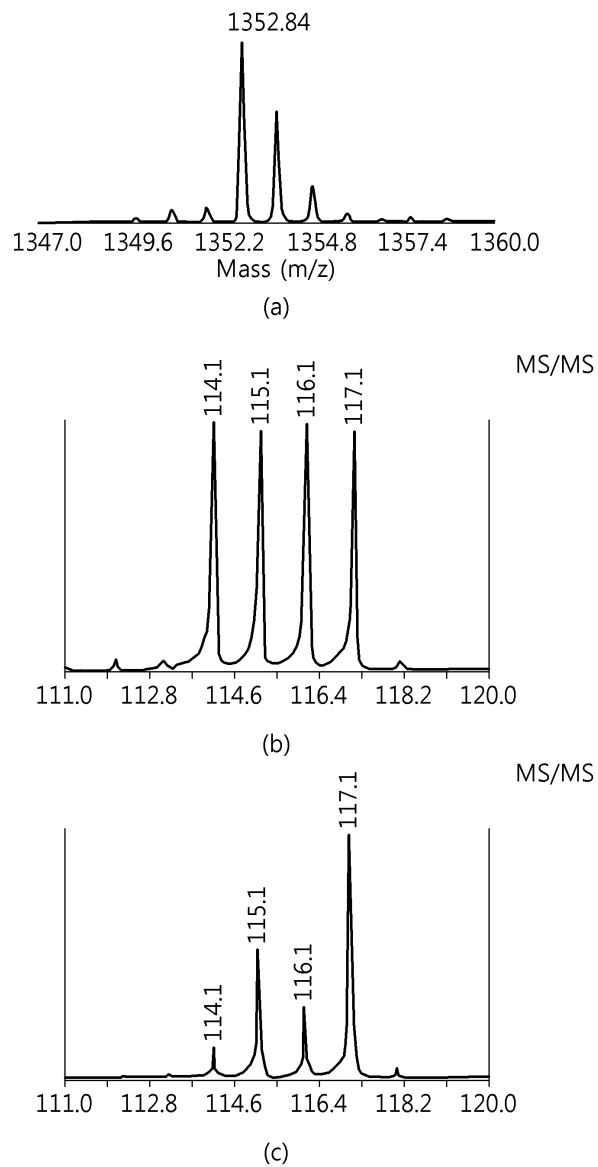
도면3



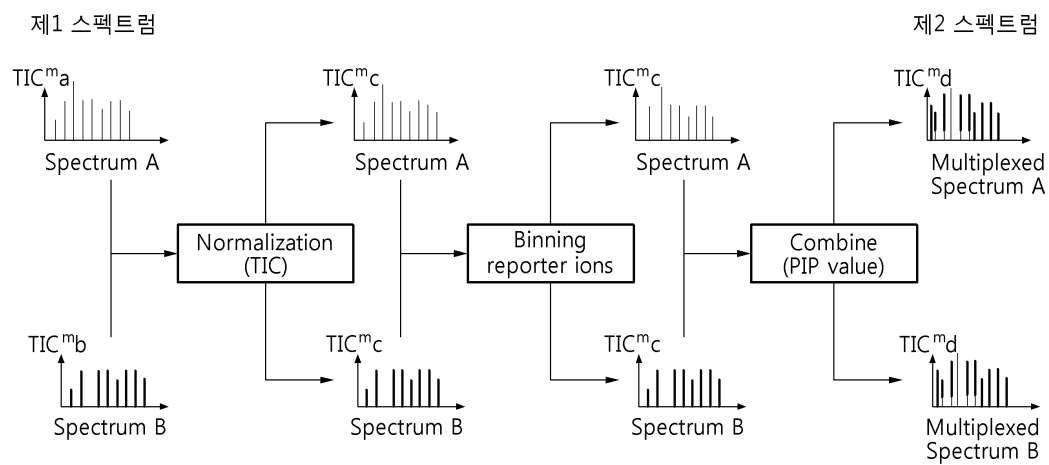
도면4



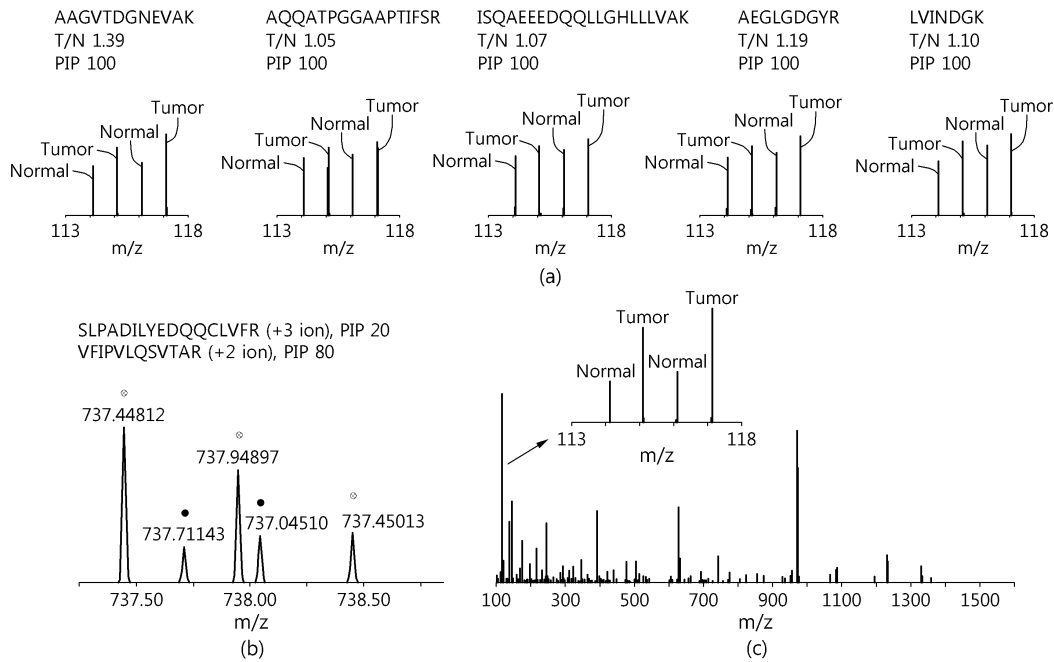
도면5



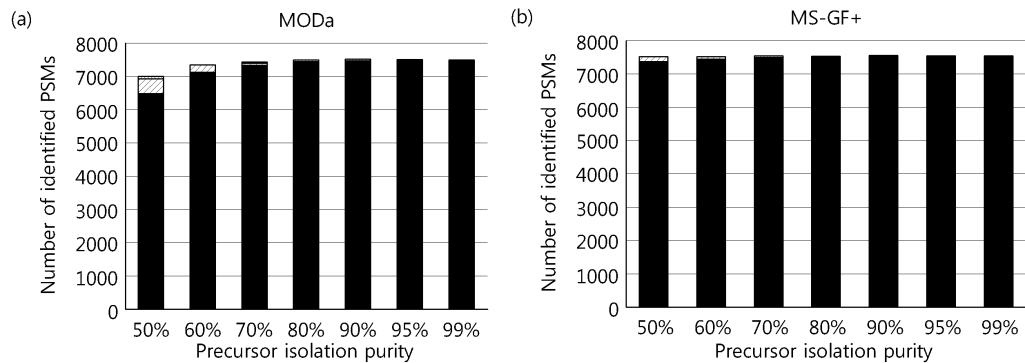
도면6



도면7



도면8

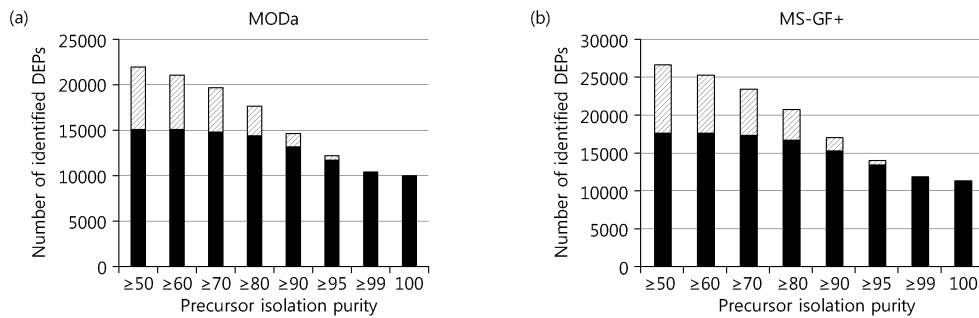


MODa	Precursor isolation purity							MS-GF+	Precursor isolation purity						
	50%	60%	70%	80%	90%	95%	99%		50%	60%	70%	80%	90%	95%	99%
■ Correct ID	6483 (86.4 %)	7130 (95.1 %)	7355 (98.1 %)	7449 (99.3 %)	7487 (99.8 %)	7497 (100.0 %)	7499 (100.0 %)	■ Correct ID	7351 (98.0 %)	7420 (98.9 %)	7461 (99.5 %)	7480 (99.7 %)	7495 (99.9 %)	7498 (100.0 %)	7498 (100.0 %)
☐ Incorrect ID(Noise)	470 (6.3 %)	189 (2.5 %)	59 (0.8 %)	16 (0.2 %)	2 (0.0 %)	0 (0.0 %)	0 (0.0 %)	☐ Incorrect ID(Noise)	136 (1.8 %)	71 (0.9 %)	33 (0.4 %)	17 (0.2 %)	4 (0.1 %)	1 (0.0 %)	1 (0.0 %)
☐ Incorrect ID(Other)	35 (0.5 %)	16 (0.2 %)	15 (0.2 %)	17 (0.2 %)	6 (0.1 %)	1 (0.0 %)	0 (0.0 %)	☐ Incorrect ID(Other)	7 (0.1 %)	4 (0.1 %)	3 (0.0 %)	2 (0.0 %)	1 (0.0 %)	1 (0.0 %)	1 (0.0 %)
Total	6988 (93.2 %)	7335 (97.8 %)	7429 (99.1 %)	7482 (99.9 %)	7495 (100.0 %)	7498 (100.0 %)	7499 (100.0 %)	Total	7494 (99.9 %)	7495 (99.9 %)	7497 (100.0 %)	7499 (100.0 %)	7500 (100.0 %)	7500 (100.0 %)	7500 (100.0 %)

도면9

Database Tools	Cutoffvalue	Precursor isolation purity						
		50%	60%	70%	80%	90%	95%	99%
MODa	MC ₁	-1.91	-1.76	-1.59	-1.41	-1.22	-1.12	-1.02
	MC ₂	2	1.79	1.63	1.45	1.25	1.13	1.03
MS-GF+	MC ₁	-1.93	-1.76	-1.58	-1.41	-1.22	-1.12	-1.02
	MC ₂	1.98	1.79	1.64	1.45	1.25	1.13	1.03
Proteome Discoverer	MC ₁	-1.98	-1.76	-1.59	-1.41	-1.22	-1.12	-1.03
	MC ₂	-2	1.79	1.64	1.45	1.25	1.13	1.03

도면10



MODa	Precursor isolation purity								MS-GF+	Precursor isolation purity							
	≥ 50%	≥ 60%	≥ 70%	≥ 80%	≥ 90%	≥ 95%	≥ 99%	100%		≥ 50%	≥ 60%	≥ 70%	≥ 80%	≥ 90%	≥ 95%	≥ 99%	100%
■ DEPs	15073 (68.7 %)	14989 (71.4 %)	14795 (75.5 %)	14336 (81.4 %)	13177 (90.1 %)	11679 (96.3 %)	10248 (99.6 %)	9891 (100.0 %)	■ DEPs	17601 (66.2 %)	17480 (69.4 %)	17220 (73.9 %)	16633 (80.4 %)	15210 (89.7 %)	13412 (96.1 %)	11711 (99.7 %)	11234 (100.0 %)
□ Putative false Positives	6858 (31.3 %)	5991 (28.6 %)	4808 (24.5 %)	3285 (18.6 %)	1446 (9.9 %)	454 (3.7 %)	37 (0.4 %)	0 (0.0 %)	□ Putative false Positives	8972 (33.8 %)	7695 (30.6 %)	6082 (26.1 %)	4044 (19.6 %)	1742 (10.3 %)	544 (3.9 %)	41 (0.3 %)	0 (0.0 %)
Total	21938 (100.0 %)	20985 (100.0 %)	19607 (100.0 %)	17621 (100.0 %)	14622 (100.0 %)	12113 (100.0 %)	10285 (100.0 %)	9891 (100.0 %)	Total	26573 (100.0 %)	25175 (100.0 %)	23302 (100.0 %)	20677 (100.0 %)	16952 (100.0 %)	13956 (100.0 %)	11752 (100.0 %)	11234 (100.0 %)