(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0127478
(43) 공개일자 2021년10월22일

(51) 국제특허분류(Int. Cl.)
G16B 40/20 (2019.01)
(52) CPC특허분류
G16B 40/20 (2019.02)
G16B 20/40 (2019.02)
(21) 출원번호 10-2020-0045382
(22) 출원일자 2020년04월14일
심사청구일자 2020년04월14일

(71) 출원인
주식회사 클리노믹스
울산광역시 울주군 언양읍 유니스트길 50, 110동 3층 301-3호
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
박종화
울산광역시 울주군 언양읍 반천강변길 51, 101동 1701호(현대아파트)
김병철
경기도 수원시 영통구 봉영로1744번길 11, 222동 802호(영통동, 벽산.풍림아파트)
(뒷면에 계속)
(74) 대리인
특허법인메이저

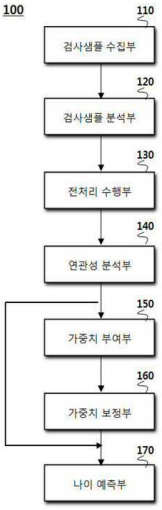
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 **다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법 및 시스템**

(57) 요 약

본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템은 피검자의 DNA, RNA 중 적어도 하나 이상을 포함하는 복수 개의 유전자 검사샘플을 수집하는 검사샘플 수집부; 복수 종류의 오믹스 데이터를 상기 복수 개의 유전자 검사샘플 각각으로부터 분석해내는 검사샘플 분석부; 상기 검사샘플 분석부를 통하여 분석된 오믹스 데이터에 대한 전처리를 수행하는 전처리 수행부; 상기 전처리 수행부를 통하여 변환된 오믹스 영역별 데이터 각각의 오믹스 종류에 기초한 연관성 분석을 수행하는 연관성 분석부; 및 상기 연관성 분석부의 분석 결과 및 상기 오믹스 영역별 데이터에 기초하여 피검자의 나이를 예측하는 나이 예측부를 포함한다.

대 표 도 - 도1



(72) 발명자

조윤성

경기도 용인시 기흥구 흥덕1로79번길 37, 503동 1702호(영덕동, 흥덕마을5단지 호반베르디움)

이황열

경기도 용인시 기흥구 흥덕1로79번길 37, 504동 1201호(영덕동, 흥덕마을5단지 호반베르디움)

명 세 서

청구범위

청구항 1

피검자의 DNA, RNA 중 적어도 하나 이상을 포함하는 복수 개의 유전자 검사샘플을 수집하는 검사샘플 수집부;

복수 종류의 오믹스 데이터를 상기 복수 개의 유전자 검사샘플 각각으로부터 분석해내는 검사샘플 분석부;

상기 검사샘플 분석부를 통하여 분석된 오믹스 데이터에 대한 전처리를 수행하는 전처리 수행부;

상기 전처리 수행부를 통하여 변환된 오믹스 영역별 데이터 각각의 오믹스 종류에 기초한 연관성 분석을 수행하는 연관성 분석부;

상기 연관성 분석부의 분석 결과 및 상기 오믹스 영역별 데이터에 기초하여 피검자의 나이를 예측하는 나이 예측부를 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템.

청구항 2

청구항 1에 있어서,

상기 복수 종류의 오믹스 데이터는 텔로미어 길이, 메틸레이션, 유전자 발현 중 적어도 하나 이상을 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템.

청구항 3

청구항 1에 있어서,

상기 전처리 수행부는

상기 복수 종류의 오믹스 데이터의 각 마커 수치를 0~1 범위의 백분위 값으로 변환하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템.

청구항 4

청구항 1에 있어서,

상기 연관성 분석부는

다중선형회귀분석, 인공신경망기반 회귀분석 중 어느 하나를 이용하여 복수의 오믹스 영역의 결정계수(R_x^2), 유의도($PVAL_x$), 평균절대오차(MAE_x) 중 적어도 하나 이상을 분석하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템.

청구항 5

청구항 1에 있어서,

상기 나이 예측부는

상기 연관성 분석부에서 분석한 복수 종류의 오믹스 영역별 분석결과 데이터를 통합(합산)하여 피검자의 나이를 예측하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템.

청구항 6

청구항 1에 있어서,

상기 연관성 분석부를 통하여 분석된 결정계수(R_x^2)에 기초하여 오믹스 데이터의 종류별로 가중치가 부여되는 가중치 부여부; 및

상기 오믹스 데이터의 종류별로 부여된 가중치를 보정하기 위한 가중치 보정부를 포함하는 다양한 오믹스 데이

터 분석에 기초한 생체나이 예측 시스템.

청구항 7

청구항 1에 있어서,

상기 연관성 분석부를 통하여 분석된 유의도($PVAL_x$)에 기초하여 오믹스 데이터의 종류별로 가중치가 부여되는 가중치 부여부; 및

상기 오믹스 데이터의 종류별로 부여된 가중치를 보정하기 위한 가중치 보정부를 더 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템.

청구항 8

청구항 1에 있어서,

상기 연관성 분석부를 통하여 분석된 평균절대오차(MAE_x)에 기초하여 오믹스 데이터의 종류별로 가중치가 부여되는 가중치 부여부; 및

상기 오믹스 데이터의 종류별로 부여된 가중치를 보정하기 위한 가중치 보정부를 더 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템.

청구항 9

청구항 6 내지 청구항 8 중 어느 한 항에 있어서,

상기 가중치 보정부를 통하여 보정된 가중치, 연관성 분석부의 분석 결과 및 상기 오믹스 영역별 데이터에 기초하여, 피검자의 나이를 상기 나이 예측부에서 아래 수학식을 이용하여 예측하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템.

[수학식]

$$AGE_{integ} = \frac{\Sigma(AGE_x * w_{x,rev})}{\Sigma w_{x,rev}}$$

생체나이(AGE_{integ}) 예측식;

청구항 10

검사샘플 수집부에서 피검자의 DNA, RNA 중 적어도 하나 이상을 포함하는 복수 개의 유전자 검사샘플을 수집하는 단계;

검사샘플 분석부에서 복수 종류의 오믹스 데이터를 상기 복수 개의 유전자 검사샘플 각각으로부터 분석하는 단계;

전처리 수행부에서 상기 검사샘플 분석부를 통하여 분석된 오믹스 데이터에 대한 전처리를 수행하는 단계;

연관성 분석부에서 상기 전처리 수행부를 통하여 변환된 오믹스 영역별 데이터 각각의 오믹스 종류에 기초한 연관성 분석을 수행하는 단계; 및

나이 예측부에서 상기 연관성 분석부의 분석 결과 및 상기 오믹스 영역별 데이터에 기초하여 피검자의 나이를 예측하는 단계를 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법.

청구항 11

청구항 10에 있어서,

상기 복수 종류의 오믹스 데이터는 텔로미어 길이, 메틸레이션, 유전자 발현 중 적어도 하나 이상을 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법.

청구항 12

청구항 10에 있어서,

상기 전처리를 수행하는 단계는

상기 복수 종류의 오믹스 데이터의 각 마커 수치를 0~1 범위의 백분위 값으로 변환하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 방법.

청구항 13

청구항 10에 있어서,

상기 연관성 분석을 수행하는 단계는

다중선형회귀분석, 인공신경망기반 회귀분석 중 어느 하나를 이용하여 복수의 오믹스 영역의 결정계수(R_x^2), 유의도($PVAL_x$), 평균절대오차(MAE_x) 중 적어도 하나 이상을 분석하는 단계인 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 방법.

청구항 14

청구항 10에 있어서,

상기 나이를 예측하는 단계는

상기 연관성 분석부에서 분석한 복수 종류의 오믹스 영역별 분석결과 데이터를 통합(합산)하여 피검자의 나이를 예측하는 단계인 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 방법.

청구항 15

청구항 10에 있어서,

가중치 부여부에서 상기 연관성 분석부를 통하여 분석된 결정계수(R_x^2)에 기초하여 오믹스 데이터의 종류별로 가중치를 부여하는 단계; 및

가중치 보정부에서 상기 오믹스 데이터의 종류별로 부여된 가중치를 보정하는 단계를 더 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 방법.

청구항 16

청구항 10에 있어서,

가중치 부여부에서 상기 연관성 분석부를 통하여 분석된 유의도($PVAL_x$)에 기초하여 오믹스 데이터의 종류별로 가중치를 부여하는 단계; 및

가중치 보정부에서 상기 오믹스 데이터의 종류별로 부여된 가중치를 보정하는 단계를 더 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 방법.

청구항 17

청구항 10에 있어서,

가중치 부여부에서 상기 연관성 분석부를 통하여 분석된 평균절대오차(MAE_x)에 기초하여 오믹스 데이터의 종류별로 가중치를 부여하는 단계; 및

가중치 보정부에서 상기 오믹스 데이터의 종류별로 부여된 가중치를 보정하는 단계를 더 포함하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 방법.

청구항 18

청구항 15 내지 청구항 17 중 어느 한 항에 있어서,

상기 가중치 보정부를 통하여 보정된 가중치, 연관성 분석부의 분석 결과 및 상기 오믹스 영역별 데이터에 기초

하여, 피검자의 나이를 상기 나이 예측부에서 아래 수학식을 이용하여 예측하는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 방법.

[수학식]

$$AGE_{integ} = \frac{\Sigma(AGE_x * w_{x,rev})}{\Sigma w_{x,rev}}$$

생체나이(AGE_{integ}) 예측식;

발명의 설명

기 술 분 야

[0001] 본 발명은 다양한 오믹스 데이터 분석에 기초한 나이 예측 방법 및 그 시스템에 관한 것으로, 구체적으로는 피검자로부터의 검체(ex. 혈액 등) 샘플에서 텔로미어 길이, DNA 메틸레이션(methylation), mRNA 발현량 등에 관한 다양한 오믹스 데이터를 획득한 후 통합적으로 분석하여 피검자의 생체나이(biological age)를 예측하고, 각 오믹스 데이터에 기반한 노화의 정도를 구분하여 분석하는 DNA 메틸레이션(methylation), mRNA 발현량, 텔로미어 길이 등의 정보를 통합적으로 활용한 노화 또는 생체나이 예측 방법 및 이를 수행하는 시스템에 관한 것이다.

배 경 기 술

[0002] 생체나이(biological age)는 전반적인 건강 상태와 노화 정도를 종합적으로 평가하여 수치화한 나이를 지칭하는 것으로, 이러한 노화/생체나이를 예측함에 있어 텔로미어 길이를 활용한 방법이 기존부터 일반적으로 이용되어 오고 있으며, 나이에 따라 유의하게 변하는 DNA 메틸레이션 또는 유전자 발현량을 기반으로 나이를 예측하기 위한 바이오마커와 이들의 조합공식 등이 개발되고 있는 추세이다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) US20030148350A1

발명의 내용

해결하려는 과제

[0004] 본 발명이 해결하고자 하는 과제는 종래의 문제점을 해결할 수 있는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법 및 시스템을 제공하는 데 있다

과제의 해결 수단

[0005] 상기 과제를 해결하기 위한 본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템은 피검자의 DNA, RNA가 포함된 복수 개의 유전자 검사샘플을 수집하는 검사샘플 수집부; 복수 종류의 오믹스 데이터를 상기 복수 개의 유전자 검사샘플 각각으로부터 분석해내는 검사샘플 분석부; 상기 검사샘플 분석부를 통하여 분석된 오믹스 데이터에 대한 전처리를 수행하는 전처리 수행부; 상기 전처리 수행부를 통하여 변환된 오믹스 영역별 데이터 각각의 오믹스 종류에 기초한 연관성 분석을 수행하는 연관성 분석부; 상기 연관성 분석부의 분석 결과 및 상기 오믹스 영역별 데이터에 기초하여 피검자의 나이를 예측하는 나이 예측부를 포함한다.

[0006] 상기 과제를 해결하기 위한 본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법은 검사샘플 수집부에서 피검자의 DNA, RNA가 포함된 복수 개의 유전자 검사샘플을 수집하는 단계; 검사샘플 분석부에서 복수 종류의 오믹스 데이터를 상기 복수 개의 유전자 검사샘플 각각으로부터 분석하는 단계; 전처리 수행부에서 상기 검사샘플 분석부를 통하여 분석된 오믹스 데이터에 대한 전처리를 수행하는 단계; 연관성 분석부에서 상기 전처리 수행부를 통하여 변환된 오믹스 영역별 데이터 각각의 오믹스 종류에 기초한 연관성 분석을

수행하는 단계; 및 나이 예측부에서 상기 연관성 분석부의 분석 결과 및 상기 오믹스 영역별 데이터에 기초하여 피검자의 나이를 예측하는 단계를 포함한다.

발명의 효과

[0007] 본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법 및 시스템을 이용하면, 생체나이 예측모델에 여러 오믹스 영역의 마커를 조합하여 반영함으로써, 개별적인 오믹스 영역에 존재하는 오차를 상쇄하고 더 정확한 나이예측이 가능하며, 통합적으로 예측된 생체나이(피검자의 현재 노화정도)에 대하여 각 오믹스 영역의 영향력(또는 노화정도)을 구분하여 해석할 수 있다는 이점이 있다.

[0008] 즉, 사람의 혈액 등의 샘플의 유전체(텔로미어 길이), 외유전체(메틸레이션), 전사체(유전자 발현) 세 가지 오믹스 데이터 등의 조합을 통해 1. 각 오믹스 데이터에 내재된 노이즈를 상쇄하여 나이예측 정확도를 높일 수 있으며, 2. 피검자의 생체 나이(노화 정도)를 오믹스 영역별로 구분하여 해석할 수 있다.

도면의 간단한 설명

[0009] 도 1은 본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측시스템의 구성도이다.

도 2a 내지 도 2d는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법을 설명한 흐름도이다.

도 3은 텔로미어 길이를 활용한 선형회귀 기반의 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 4는 메틸레이션 마커 16개를 활용한 선형회귀 기반 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 5는 유전자발현 마커 18개를 활용한 선형회귀 기반 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 6은 본 발명에서 제시한 텔로미어 길이, 메틸레이션 마커, 유전자발현 마커를 조합한 선형회귀 기반 오믹스 통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 7은 본 발명에서 제시한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치(결정계수 가중)로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 8은 메틸레이션 마커 4개를 활용한 선형회귀 기반 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 9은 유전자발현 마커 4개를 활용한 선형회귀 기반 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 10은 본 발명에서 제시한 텔로미어 길이, 메틸레이션, 유전자발현 마커를 조합한 선형회귀 기반 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 11은 본 발명에서 제시한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치(결정계수 가중)로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 12는 텔로미어 길이를 활용한 인공신경망 기반의 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 13은 메틸레이션 마커 16개를 활용한 인공신경망 기반의 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 14는 유전자발현 마커 18개를 활용한 인공신경망 기반의 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 15는 본 발명에서 제시한 텔로미어 길이, 메틸레이션 마커, 유전자발현 마커를 조합한 인공신경망 기반 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 16은 본 발명에서 제시한 인공신경망을 이용한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치(결정계수 가중)로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 17은 본 발명에서 제시한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치(유의도 가중)로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

도 18은 본 발명에서 제시한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역

별 가중치(평균오차 가중)로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0010] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0011] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미하며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0012] 명세서 전체에서 사용되는 정도의 용어 "약", "실질적으로" 등은 언급된 의미에 고유한 제조 및 물질 허용오차가 제시될 때 그 수치에서 또는 그 수치에 근접한 의미로 사용되고, 본 발명의 이해를 돕기 위해 정확하거나 절대적인 수치가 언급된 개시 내용을 비양심적인 침해자가 부당하게 이용하는 것을 방지하기 위해 사용된다. 본 발명의 명세서 전체에서 사용되는 정도의 용어 "~(하는) 단계" 또는 "~의 단계"는 "~ 를 위한 단계"를 의미하지 않는다.

[0013] 본 명세서에 있어서 '부(部)'란, 하드웨어에 의해 실현되는 유닛(unit), 소프트웨어에 의해 실현되는 유닛, 양방을 이용하여 실현되는 유닛을 포함한다. 또한, 1개의 유닛이 2개 이상의 하드웨어를 이용하여 실현되어도 되고, 2개 이상의 유닛이 1개의 하드웨어에 의해 실현되어도 된다.

[0014] 본 명세서에 있어서 단말, 장치 또는 디바이스가 수행하는 것으로 기술된 동작이나 기능 중 일부는 해당 단말, 장치 또는 디바이스와 연결된 서버에서 대신 수행될 수도 있다. 이와 마찬가지로, 서버가 수행하는 것으로 기술된 동작이나 기능 중 일부도 해당 서버와 연결된 단말, 장치 또는 디바이스에서 수행될 수도 있다.

[0015] 본 명세서에서 있어서, 단말과 매핑(Mapping) 또는 매칭(Matching)으로 기술된 동작이나 기능 중 일부는, 단말의 식별 정보(Identifying Data)인 단말기의 고유번호나 개인의 식별정보를 매핑 또는 매칭한다는 의미로 해석될 수 있다.

[0016] 이하 첨부된 도면을 참고하여 본 발명을 상세히 설명하기로 한다.

[0017] 도 1은 본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측시스템의 구성도이고, 도 2a 내지 도 2d는 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법을 설명한 흐름도이고, 도 3은 텔로미어 길이를 활용한 선형회귀 기반의 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 4는 메틸레이션 마커 16개를 활용한 선형회귀 기반 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 5는 유전자발현 마커 18개를 활용한 선형회귀 기반 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 6은 본 발명에서 제시한 텔로미어 길이, 메틸레이션 마커, 유전자발현 마커를 조합한 선형회귀 기반 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 7은 본 발명에서 제시한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 8은 메틸레이션 마커 4개를 활용한 선형회귀 기반 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 9은 유전자발현 마커 4개를 활용한 선형회귀 기반 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 10은 본 발명에서 제시한 텔로미어 길이, 메틸레이션, 유전자발현 마커를 조합한 선형회귀 기반 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 11은 본 발명에서 제시한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치(결정계수 가중)로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 12는 텔로미어 길이를 활용한 인공신경망 기반의 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 13은 메틸레이션 마커 16개를 활용한 인공신경망 기반의 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 14는 유전자발현 마커 18개를 활용한 인공신경망 기반의 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 15는 본 발명에서 제시한 텔로미어 길이, 메틸레이션 마커, 유전자발현 마커를 조합한 인공신경망 기반 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 16은 본 발명에서 제시한 인공신경망을 이용한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치(결정계수 가중)로 합

산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 17은 본 발명에서 제시한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치(유의도 가중)로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이고, 도 18은 본 발명에서 제시한 텔로미어 기반 나이, 메틸레이션 기반 나이, 유전자발현 기반 나이를 오믹스 영역별 가중치(평균오차 가중)로 합산한 오믹스통합 생체나이와 실제나이 간의 상관관계를 나타낸 그래프이다.

[0018] 먼저, 도 1에 도시된 바와 같이, 본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측시스템(100)은 검사샘플(ex. 혈액 등)로부터 다양한 오믹스 정보(ex. 텔로미어 길이, DNA메틸레이션 정도, 유전자 발현량 등..)를 획득하고, 각각의 획득된 오믹스 정보들을 통합적으로 분석(ex. 다중 선형회귀분석, 오믹스 영역별 가중치 부여)하여 생체나이를 기존대비 더욱 정확하게 측정(예측)할 수 있는 시스템이다.

[0019] 보다 구체적으로, 본 발명의 다양한 오믹스 데이터 분석에 기초한 생체나이 예측 시스템(100)은 검사샘플 수집부(110), 검사샘플 분석부(120), 전처리 수행부(130) 연관성 분석부(140), 가중치 부여부(150), 가중치 보정부(160) 및 나이 예측부(170)를 포함한다.

[0020] 상기 검사샘플 수집부(110)는 피검자의 DNA, RNA가 포함된 복수 개의 유전자 검사샘플을 수집하는 구성으로, 즉, 복수 개의 노화 바이오마커 검사샘플을 수집한 후, 분류하는 구성이다.

[0021] 여기서, 노화 바이오마커는 다양한 나이대의 혈액 샘플로부터 DNA를 채취하여 텔로미어 길이를 측정한 결과 및 Methyl-seq 또는 chip 실험을 통해 DMR 분석 등을 수행한 결과, RNA를 채취하여 RNA-seq 또는 microarray 실험을 통해 DEG 분석 등을 수행한 결과정보일 수 있다.

[0022] 또한, 각 노화 바이오마커 검사샘플은 학습 데이터 및 테스트 데이터로 분류되고, 분류된 학습 데이터 및 테스트 데이터는 각 오믹스 영역 마커들의 나이 예측과 통합오믹스 분석 및 예측나이 가중치 합산 분석 시에 이용된다.

[0023] 다음으로, 검사샘플 분석부(120)는 복수 종류의 오믹스 데이터를 상기 복수 개의 유전자 검사샘플 각각으로부터 분석해내는 구성으로, 즉, 상기 복수 개의 유전자 검사샘플로부터 각 유전자 별 텔로미어 길이정보, DNA 메틸레이션 정보, 유전자 발현량을 포함하는 오믹스 영역을 분석하는 구성일 수 있다.

[0024] 보다 구체적으로, 상기 검사샘플 분석부(120)는 텔로미어 길이 측정부, 메틸레이션 마커 분석 및 필터링부 및 유전자발현 마커 분석 및 필터링부를 포함한다.

[0025] 상기 텔로미어 길이 측정부는 qPCR, TRF 또는 Q-FISH 방식 등을 이용하여 단일복제유전자(single copy gene) 대비 텔로미어의 상대적 길이를 측정하는 구성으로, 길이가 알려진 표준 올리고머 샘플로부터 농도별 형광검출한계 사이클 수(Ct)를 측정한 후, 텔로미어의 Ct값을 기준유전자의 Ct값으로 나누어 전체 텔로미어 길이를 구하고, 이를 인간 유전체 텔로미어 수로 나누어 텔로미어의 절대길이를 측정한다.

[0026] 참고로, 상술한 텔로미어 길이 측정법은 일 예에 불과할 뿐, 일반적인 기존의 다양한 텔로미어 길이 측정법이 적용될 수 있다.

[0027] 다음으로, 메틸레이션 마커 분석 및 필터링부는 Methyl-seq, chip 등 실험을 통해 DMR 분석 등을 이용하여 획득한 메틸레이션 로우 데이터(raw data)를 인간게놈지도(human reference)에 맵핑(mapping)하여 각 검사샘플의 위치별 메틸레이션(methylation) 정도(이하, 베타 값(beta value))를 구하고, DMR 분석 등을 이용하여 각 검사샘플에서 나이에 따라 베타 값이 증가하거나 감소하는 영역을 선별한다.

[0028] 다음으로, 유전자발현 마커 분석 및 필터링부는 RNA-seq, microarray 등 실험을 통해 획득한 유전자발현 로우 데이터(raw data)를 인간게놈지도(human reference)에 맵핑(mapping)하여 각 검사샘플의 유전자별 발현량을 산출하고, 산출된 유전자 발현량에서 성별/생활습관 등에 따른 배치 효과(batch effect)를 제거한 후, DEG 분석 등을 이용하여 각 검사 샘플에서 나이에 따라 발현량이 증가하거나 감소하는 유전자를 선별한다.

[0029] 다음으로, 상기 전처리 수행부(130)는 검사샘플 분석부(120)를 통하여 분석된 오믹스 데이터에 대한 전처리를 수행하는 구성이다.

[0030] 보다 구체적으로, 전처리 수행부(130)는 선정된 메틸레이션 마커 및 유전자발현 마커의 베타 값과 발현량 값 및 텔로미어 길이를 다중선형회귀분석 또는 인공신경망 기반 회귀분석에 적용하기 위하여 0~1 범위의 백분위 값으로 변환한다.

[0031] 다음으로, 상기 연관성 분석부(140)는 상기 전처리 수행부(130)를 통해 변환된 오믹스 영역별 데이터 각각의 오

믹스 종류에 기초한 연관성 분석을 수행한다. 보다 구체적으로, 연관성 분석부(140)는 다중선형회귀분석 또는 인공신경망 기반 회귀분석을 이용하여 변환된 오믹스 영역별 바이오마커의 전처리값을 독립변수로하고 생체나이를 종속변수로하는 회귀모델에서 독립변수의 각 계수값을 산출한다. 산출된 계수를 통해 각 오믹스 영역 바이오마커의 전처리값으로부터 예측된 영역별 생체나이와 실제나이 간의 연관성을 분석하고, 분석된 연관성은 결정계수(R_x^2), 유의도($PVAL_x$), 평균절대오차(MAE_x) 중 어느 하나 일 수 있다.

[0032] 다음으로, 가중치 부여부(150)는 연관성 분석부를 통해 분석된 연관성(결정계수, 유의도, 평균절대오차) 중 어느 하나에 기초하여 오믹스 데이터의 종류별로 가중치를 부여하는 구성일 수 있다.

[0033] 상기 가중치 부여부(150)는 각 오믹스 영역의 결정계수(R_x^2), 유의도($PVAL_x$), 평균절대오차(MAE_x)에 대한 가중치(W_x)를 하기의 식들을 통해 산출한다.

$$W_x = 1/(1-R_x^2)$$

[0034] ……..(결정계수 관련 가중치 식)

$$W_x = \log(PVAL_x) * (-1)$$

[0035] ……(유의도 관련 가중치 식)

$$W_x = 1/mae_{x,rev}$$

[0036] …….(평균절대오차 관련 가중치 식)

[0038] 상기 가중치 보정부(160)는 하기의 식들을 이용하여 상기 오믹스 데이터의 종류별로 부여된 가중치의 영역별 가중치 평균값(W_{avg})의 지수화를 통해 보정 가중치($w_{x,rev}$)를 구하는 구성일 수 있다.

[0039] $$w_{x,rev} = W_{avg}^{(W_x/W_{avg})}$$ …….(가중치 보정식)

[0040] 한편, 상기 가중치 보정부(160)는 각 오믹스 영역의 평균절대오차(MAE_x)에 대한 가중치 보정 전, 실제 나이 분포 대비 평균절대오차를 상대적으로 반영하기 위해 샘플그룹의 평균 나이(AGE_{avg})에 의한 분포보정($mae_{x,rev}$)을 하기의 식을 통해 구할 수 있다.

$$mae_{x,rev} = MAE_x/AGE_{avg}$$

[0041]

[0042] 다음으로, 나이 예측부(170)는 상기 연관성 분석부의 분석결과 및 상기 오믹스 영역별 데이터에 기초하여 피검자의 나이를 예측하는 구성으로, 하기의 식을 통해 피검자의 나이를 예측할 수 있다.

$$AGE_{integ} = \frac{\sum(AGE_x * w_{x,rev})}{\sum w_{x,rev}}$$

[0043] …….(생체나이 AGE_{integ} 예측식)

[0044] 즉, 나이 예측부(170)는 개별 오믹스 데이터의 나이에 대한 결정계수, 유의도, 평균절대오차 중 어느 하나를 이용하여 각 오믹스 영역의 가중치를 점수화한 후, 개별 오믹스로부터 추론된 나이를 가중치에 따라 합산한 합산값을 비교하여 생체 나이 또는 노화 상태를 예측하는 구성일 수 있다.

[0046] 이하에서는 도면을 참조하여, 본원에 개시된 구성들을 통해 텔로미어 길이, 메틸레이션 마커 16개, 유전자발현 마커 18개를 활용한 선형회귀기반 생체나이의 실제나이에 대한 예측력을 비교한 제1 비교예를 간략하게 설명하

도록 한다.

[0047] 1) 오믹스통합 다중선형회귀분석

[0048] 제1 비교예에서 본원의 연관성 분석부(140)는 미리 선정한 보정 유의도(adjusted p-value) < 1.0e-30 기준의 메틸레이션 마커 16개와 adj.pval < 5.0e-02 기준의 유전자발현 마커 18개를 이용하여 텔로미어 길이와 함께 각 영역별 다중선형회귀분석 및 오믹스 통합 분석을 수행한다.

[0049] 2) 오믹스 영역별 생체나이 가중치 합산분석(결정계수 가중)

[0050] 본원의 연관성 분석부(140)는 각 오믹스 영역의 마커를 활용한 다중선형회귀분석으로부터 영역별로 예측된 생체나이의 샘플 실제나이에 대한 결정계수(R_x^2)를 구한다.

[0051] 본원의 가중치 부여부(150)는 결정계수가 큰 오믹스 영역에 더 큰 가중치를 주기 위해, 각 오믹스 영역의 가중치(W_x)를 식 1과 같이 산출한다.

[0052] [식 1]

$$W_x = 1/(1-R_x^2)$$

[0053]

[0054] 또한, 각 오믹스 영역간 생체나이의 실제 나이에 대한 결정계수 차이가 큰 경우, 본원의 가중치 보정부(160)는 신뢰도가 높은 오믹스 영역의 나이를 영역별 가중치(W_x)에 강조하여 반영하기 위해 각 오믹스 영역 가중치 평균값(W_{avg})의 지수화를 통해 보정된 가중치($w_{x,rev}$)를 식 2와 같이 산출한다.

[0055] 본원의 나이 예측부(170)는 식 3과 같이 영역별로 예측된 나이(AGE_x)에 오믹스 영역별 가중치를 적용하여 합산함으로써 오믹스통합 생체나이(AGE_{integ})를 계산한다.

[0056] [식 2]

$$w_{x,rev} = W_{avg}^{(W_x/W_{avg})}$$

[0057]

[0059] [식 3]

$$AGE_{integ} = \frac{\sum(AGE_x * w_{x,rev})}{\sum w_{x,rev}}$$

[0060]

[0061] 하기의 표 1은 각 오믹스 영역의 결정계수, 가중치, 보정 가중치를 나타낸 표이고, 표 2는 각 오믹스 영역별 나이 및 각 오믹스 영역별 가중치를 합산하여 오믹스통합 생체나이를 예측한 값을 실제나이와 비교한 표이다.

표 1

[0063]

	텔로미어	메틸레이션	유전자발현
결정계수(R_x^2)	0.317	0.930	0.834
가중치 W_x	1.46	14.25	6.02
보정가중치 $w_{x,rev}$	1.49	49.19	5.19

표 2

[0064]

AGE_{telo}	w_{telo}	AGE_{meth}	w_{meth}	AGE_{exp}	w_{exp}	AGE_{integ}	실제나이
38.63	1.49	21.31	49.19	26.46	5.19	22.25	22
34.38		46.24		47.15		46.00	44
57.15		66.77		67.77		66.61	74

[0066] 표 3은 개별오믹스 생체나이예측 대비 오믹스통합 회귀분석 생체나이 및 오믹스 영역별 나이 가중합산 오믹스통합 생체나이 예측결과를 비교한 표이다.

[0068] [표 3]

[0069]

	텔로미어 기반 나이예측	메틸레이션 기반 나이예측	유전자발현 기반 나이예측	통합오믹스 기반 나이예측	오믹스별 나이 가중치합산 결과
결정계수(R^2)	0.317	0.930	0.834	0.979	0.936
유의도(P-val)	1.7E-05	6.4E-30	9.9E-21	6.5E-43	7.2E-31
평균절대오차(MAE)	10.593	3.003	5.094	1.773	2.934

[0071] 표 3을 참조하면, 개별 오믹스로부터 다중선형회귀분석을 통해 예측된 나이에 비해 오믹스통합 회귀분석 또는 오믹스 영역별 나이 가중치합산을 통해 예측된 오믹스통합 생체나이가 결정계수 및 유의도 관점에서 샘플의 실제 나이에 가까우며, 평균오차(MAE) 가 적다는 적을 확인할 수 있다.

[0073] 이하에서는 도면을 참조하여, 본원에 개시된 구성들을 통해 텔로미어 길이, 메틸레이션 마커 4개, 유전자발현 마커 4개를 활용한 선형회귀기반 생체나이의 실제나이에 대한 예측력을 비교한 제2 비교 예를 간략하게 설명하도록 한다.

[0074] 1) 오믹스통합 다중선형회귀분석

[0075] 제2 비교예에서 본원의 연관성 분석부(140)는 adj.Pval < 1.0e-30, 그리고 마커와 실제나이간 연관성의 절대값 | R | > 0.75 기준의 메틸레이션 마커 4개와 adj.Pval < 1.0e-04 기준의 유전자발현 마커 4개를 선정하여, 텔로미어 길이와 함께 각 영역별 다중선형회귀분석 및 오믹스 통합 분석을 수행한다.

[0077] 2) 오믹스 영역별 생체나이 가중치 합산분석 (결정계수 가중)

[0078] 제1 비교 예와 동일한 방식으로 적용한다. 하기의 표 4는 각 오믹스 영역의 결정계수, 가중치, 보정 가중치를 나타낸 표이고, 표 5는 각 오믹스 영역별 나이 및 각 오믹스 영역별 가중치를 합산하여 오믹스통합 생체나이를 예측한 값을 실제나이와 비교한 표이다.

표 4

[0079]

	텔로미어	메틸레이션	유전자발현
결정계수(R_x^2)	0.310	0.860	0.717
가중치 W_x	1.46	7.14	3.54
보정가중치 $w_{x,rev}$	1.66	11.78	3.39

표 5

[0080]

AGE_{telo}	w_{telo}	AGE_{meth}	w_{meth}	AGE_{exp}	w_{exp}	AGE_{integ}	실제나이
38.63	1.66	19.04	11.78	27.73	3.39	23.94	22
34.38		45.20		45.58		44.01	44
36.77		63.13		57.61		58.34	74

[0082] 하기의 표 6은 개별 오믹스 생체나이예측 대비 오믹스통합 회귀분석 생체나이 및 나이 가중합산 오믹스통합 생체나이 예측결과를 비교한 표이다.

[0083] [표 6]

[0084]

	텔로미어 기반 나이예측	메틸레이션 기반 나이예측	유전자발현 기반 나이예측	통합오믹스 기반 나이예측	오믹스별 나이 가중치합산 결과
결정계수(R^2)	0.317	0.860	0.717	0.887	0.877
유의도(P-val)	1.7E-05	1.5E-22	4.8E-15	8.2E-25	6.4E-24
평균절대오차(MAE)	10.593	4.637	6.263	4.052	4.506

[0085] 표 6을 참조하면, 개별 오믹스로부터 다중선형회귀분석을 통해 예측된 나이에 비해 오믹스통합 회귀분석 또는 오믹스 영역별 나이 가중치합산을 통해 예측된 오믹스통합 생체나이가 결정계수 및 유의도 관점에서 샘플의 실제 나이에 가까우며, 평균오차가 적다는 것을 확인할 수 있다.

[0087] 이하에서는 도면을 참조하여, 본원에 개시된 구성들을 통해 텔로미어 길이, 메틸레이션 마커 16개, 유전자발현 마커 18개를 활용한 인공신경망기반 생체나이의 실제나이에 대한 예측력을 비교한 제3 비교예를 설명하도록 한다.

[0088] 1) 오믹스 통합 인공신경망 기반 회귀분석

[0089] 제3 비교예에서 본원의 연관성 분석부(140)는 adj.Pval<1.0e-30 기준의 메틸레이션 마커 16개와 adj.Pval<5.0e-02 기준의 유전자발현 마커 18개를 선정하여, 텔로미어 길이와 함께 각 영역별 인공신경망기반 회귀분석 및 오믹스 통합 분석을 수행한다.

[0090] 2) 오믹스 영역별 생체나이 가중치 합산분석 (결정계수 가중)

[0091] 비교 예 1과 동일한 방식으로 적용한다.

[0092] 하기의 표 7은 각 오믹스 영역의 결정계수, 가중치, 보정 가중치를 나타낸 표이고, 표 8은 각 오믹스 영역별 나이 및 각 오믹스 영역별 가중치를 합산하여 오믹스통합 생체나이를 예측한 값을 실제나이와 비교한 표이다.

표 7

[0093]

	텔로미어	메틸레이션	유전자발현
결정계수(R_x^2)	0.309	0.969	0.959
가중치 W_x	1.45	32.40	24.45
보정가중치 $w_{x,rev}$	1.25	140.86	41.82

표 8

[0094]

AGE_{telo}	w_{telo}	AGE_{meth}	w_{meth}	AGE_{exp}	w_{exp}	AGE_{integ}	실제나이
41.29	1.25	21.18	140.86	19.62	41.82	20.96	22
34.46		46.31		47.53		46.51	44
53.73		70.24		67.73		69.56	74

[0095] 표 9는 인공신경망기반 개별오믹스 생체나이예측 대비 오믹스통합 회귀분석 및 나이 가중합산 오믹스통합 생체나이 예측결과를 비교한 표이다.

[0096] [표 9]

[0097]

	텔로미어 기반 나이예측	메틸레이션 기반 나이예측	유전자발현 기반 나이예측	통합오믹스 기반 나이예측	오믹스별 나이 가중치합산 결과
결정계수(R^2)	0.309	0.969	0.959	0.979	0.971
유의도(P-val)	2.3E-05	1.1E-38	1.1E-35	6.5E-43	3.0E-39
평균절대오차(MAE)	10.656	1.712	2.285	1.773	1.671

[0098] 표 9를 참조하면, 개별 오믹스로부터 인공신경망기반 회귀분석을 통해 예측된 나이에 비해 오믹스통합 회귀분석 또는 오믹스 영역별 나이 가중치합산을 통해 예측된 오믹스통합 생체나이가 결정계수 및 유의도 관점에서 샘플의 실제 나이에 가까우며, 평균오차가 적다는 것을 알 수 있다.

[0100] 이하에서는 도면을 참조하여, 본원에 개시된 구성들을 통해 텔로미어 길이, 메틸레이션 마커 16개, 유전자발현 마커 18개를 활용한 선형회귀기반 나이예측(가중치 scoring)을 비교한 제4 비교예를 설명하도록 한다.

[0101] 1) 오믹스 통합 다중선형회귀분석

[0102] 제4 비교예에서 본원의 연관성 분석부(140)는 보정 유의도(adjusted p-value) < 1.0e-30 기준의 메틸레이션 마커 16개와 adj.Pval < 5.0e-02 기준의 유전자발현 마커 18개를 선정하여, 텔로미어 길이와 함께 각 영역별 다중선형회귀분석 및 오믹스 통합 분석을 수행한다.

[0103] 2-1) 오믹스 영역별 생체나이 가중치 합산분석(유의도 가중)

[0104] 본원의 연관성 분석부(140)는 각 오믹스 영역의 마커를 활용한 다중선형회귀분석으로부터 영역(x)별로 예측된 생체나이와 샘플 실제나이간 상관성의 유의도($PVAL_x$)를 구한다.

[0105] 본원의 가중치 부여부(150)는 매우 작은 실수범위에 분포하는 유의도의 스케일을 변환하기 위해 가중치(W_x)를 식 4와 같이 산출한다.

[0106] [식 4]

$$W_x = \log(PVAL_x) * (-1)$$

[0107]

[0108] 또한, 각 오믹스 영역간 생체나이와 실제나이의 상관성 유의도 차이가 큰 경우, 본원의 가중치 보정부(160)는 신뢰도가 높은 오믹스 영역의 나이를 영역별 가중치($W_x W_x$)에 강조하여 반영하기 위해 각 오믹스 영역 가중치 평균값(W_{avg})의 지수화를 통해 보정된 가중치($w_{x,rev}$)를 식 5와 같이 산출한다.

[0109] 본원의 나이 예측부(170)는 오믹스 영역별 가중치를 합산하여 생체나이(AGE_{integ})를 산출한다.

[0110] [식 5]

$$w_{x,rev} = W_{avg}^{(W_x/W_{avg})}$$

[0111]

[0112] [식 6]

$$AGE_{integ} = \frac{\sum(AGE_x * w_{x,rev})}{\sum w_{x,rev}}$$

[0113]

[0114] 하기의 표 10은 각 오믹스 영역의 유의도, 가중치, 보정 가중치를 나타낸 표이고, 표 11은 각 오믹스 영역별 나이 및 각 오믹스 영역별 가중치를 합산하여 오믹스통합 생체나이를 예측한 값을 실제나이와 비교한 표이다.

표 10

[0115]

	텔로미어	메틸레이션	유전자발현
유의도($PVAL_x$)	1.7E-05	6.4E-30	9.9E-21
가중치 W_x	4.77	29.20	20.01
보정가중치 $w_{x,rev}$	2.15	108.67	24.84

표 11

[0116]

AGE_{telo}	w_{telo}	AGE_{meth}	w_{meth}	AGE_{exp}	w_{exp}	AGE_{integ}	실제나이
38.63	2.15	21.31	108.67	26.46	24.84	22.53	22
34.38		46.24		47.15		46.00	44
57.15		66.77		67.77		66.81	74

[0118] 2-2) 오믹스 영역별 생체나이 가중치 합산분석 (평균오차 가중)

[0119] 본원의 연관성 분석부(140)는 각 오믹스 영역의 마커를 활용한 다중선형회귀분석으로부터 영역(x)별로 예측된 생체나이와 샘플 실제나이의 평균절대오차(MAE_x)를 구한다.

[0120] 본원의 가중치 부여부(150)는 평균절대오차가 작은 오믹스 영역에 더 큰 가중치를 주기 위해, 각 오믹스 영역의 가중치(W_x)를 식 7과 같이 산출한다.

[0121] [식 7]

$$W_x = 1/mae_{x,rev}$$

[0122]

[0123] 또한, 실제 나이 분포 대비 평균절대오차를 상대적으로 반영하기 위해 샘플그룹의 평균 나이(AGE_{avg})에 의한 분포보정($mae_{x,rev}$)을 하기의 식 8과 같이 적용하고, 각 오믹스 영역간 생체나이와 실제 나이의 평균절대오차의 차이가 큰 경우, 본원의 가중치 보정부(160)는 신뢰도가 높은 오믹스 영역의 나이를 영역별 가중치(W_x)에 강조하여 반영하기 위해 각 오믹스 영역 가중치 평균값(W_{avg})의 지수화를 통해 보정된 가중치($w_{x,rev}$)를 하기의 식 9와 같이 산출한 후, 식 10을 이용하여 오믹스 영역별 보정가중치를 합산하여 통합 생체 나이(AGE_{integ})를 계산한다.

[0124] [식 8]

$$mae_{x,rev} = MAE_x / AGE_{avg}$$

[0125]

[0126] [식 9]

$$w_{x,rev} = W_{avg}^{(w_x/w_{avg})}$$

[0127]

[0128] [식 10]

$$AGE_{integ} = \frac{\Sigma(AGE_x * w_{x,rev})}{\Sigma w_{x,rev}}$$

[0129]

[0130] 하기의 표 12는 각 오믹스 영역의 평균절대오차, 보정평균절대오차, 가중치, 보정 가중치를 나타낸 표이고, 표 13은 각 오믹스 영역별 나이 및 각 오믹스 영역별 가중치를 합산하여 오믹스통합 생체나이를 예측한 값을 실제 나이와 비교한 표이다.

표 12

[0131]

	텔로미어	메틸레이션	유전자발현
평균절대오차(MAE_x)	10.593	3.003	5.094
보정평균절대오차($mae_{x,rev}$)	0.237	0.067	0.114
가중치 W_x	4.22	14.89	8.78
보정가중치 $w_{x,rev}$	2.75	35.57	8.21

표 13

[0132]

AGE_{telo}	w_{telo}	AGE_{meth}	w_{meth}	AGE_{exp}	w_{exp}	AGE_{integ}	실제나이
38.63	2.75	21.31	35.57	26.46	8.21	23.24	22
34.38		46.24		47.15		45.70	44
57.15		66.77		67.77		66.38	74

[0133] 하기의 표 14는 개별오믹스 생체나이예측 대비 각 가중치 부여방법을 적용한 나이 가중합산 오믹스통합 생체나이 예측결과를 비교한 표이다.

[0134] [표 14]

[0135]

	텔로미어 기반 나이예측	메틸레이션 기반 나이예측	유전자발현 기반 나이예측	가중치 합산나이 (유의도 가중)	가중치 합산나이 (평균오차 가중)
결정계수(R^2)	0.317	0.930	0.834	0.938	0.938
유의도(P-val)	1.7E-05	6.4E-30	9.9E-21	3.3E-31	2.9E-31
평균절대오차(MAE)	10.593	3.003	5.094	2.991	2.987

[0137] 표 14를 참조하면, 개별 오믹스로부터 회귀분석을 통해 예측된 나이에 비해 유의도 또는 평균오차를 점수화(scoring)하여 가중치 합산한 오믹스통합 생체나이가 결정계수 및 유의도 관점에서 샘플의 실제 나이에 가까우며, 평균오차가 적다는 것을 확인할 수 있다.

[0139] 이하에서는 도 2a 및 도 2b를 참조하여, 본 발명의 제1 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법을 설명하도록 한다.

[0140] 본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법(S700)은 검사샘플 수집부(110)에서 피검자의 DNA, RNA가 포함된 복수 개의 유전자 검사샘플을 수집(S710)한 후, 검사샘플 분석부(120)에

서 복수 종류의 오믹스 데이터(텔로미어 길이, 메틸레이션, 유전자 발현 중 적어도 하나 이상을 포함)를 상기 복수 개의 유전자 검사샘플 각각으로부터 분석(S720)한 후, 전처리 수행부(130)에서 상기 검사샘플 분석부(120)를 통하여 분석된 오믹스 데이터의 각 마커 수치를 0~1 범위의 백분위 값으로 변환하는 전처리를 수행(S730)한다.

[0141] 이후, 연관성 분석부(140)에서 상기 전처리 수행부(130)를 통하여 변환된 오믹스 영역별 데이터 각각의 오믹스 종류에 기초한 연관성 분석(S740)을 수행한다.

[0142] 상기 S740 과정은 다중선형회귀분석 또는 인공신경망 기반 회귀분석을 이용하여 복수의 오믹스 영역별 데이터 간의 연관성을 분석하는 과정으로, 분석된 연관성은 결정계수(R_x^2), 유의도($PVAL_x$), 평균절대오차(MAE_x) 중 어느 하나 일 수 있다.

[0143] 이후, 나이 예측부(170)에서 상기 연관성 분석부(140)의 분석 결과 및 상기 오믹스 영역별 데이터에 기초하여 피검자의 나이를 예측(S750)한다.

[0144] S750 과정은 분석한 복수 종류의 오믹스 영역별 분석결과 데이터를 통합(합산)하여 피검자의 나이를 예측하는 과정일 수 있다.

[0146] 이하에서는 도 2c 및 도 2d를 참조하여, 본 발명의 제2 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법을 설명하도록 한다.

[0147] 본 발명의 일 실시예에 따른 다양한 오믹스 데이터 분석에 기초한 생체나이 예측방법(S800)은 검사샘플 수집부(110)에서 피검자의 유전자 검사샘플이 포함된 복수 개의 유전자 검사샘플을 수집(S810)한 후, 검사샘플 분석부(120)에서 복수 종류의 오믹스 데이터(텔로미어 길이, 메틸레이션, 유전자 발현 중 적어도 하나 이상을 포함)를 상기 복수 개의 유전자 검사샘플 각각으로부터 분석(S820)한 후, 전처리 수행부(130)에서 상기 검사샘플 분석부(120)를 통하여 분석된 오믹스 데이터의 각 마커 수치를 0~1 범위의 백분위 값으로 변환하는 전처리를 수행(S830)한다.

[0148] 이후, 연관성 분석부(140)에서 상기 전처리 수행부(130)를 통하여 변환된 오믹스 영역별 데이터 각각의 오믹스 종류에 기초한 연관성 분석(S840)을 수행한다.

[0149] 상기 S840 과정은 다중선형회귀분석 또는 인공신경망 기반 회귀분석을 이용하여 복수의 오믹스 영역별 데이터 간의 연관성을 분석하는 과정으로, 분석된 연관성은 결정계수(R_x^2), 유의도($PVAL_x$), 평균절대오차(MAE_x) 중 어느 하나 일 수 있다.

[0150] S840 과정이 완료되면, 가중치 부여부(150)에서 연관성 분석부를 통해 분석된 연관성(결정계수, 유의도, 평균절대오차) 중 어느 하나에 기초하여 오믹스 데이터의 종류별로 가중치를 부여(S850)한다.

[0151] 이때, 상기 가중치 부여부(150)는 각 오믹스 영역의 결정계수(R_x^2), 유의도($PVAL_x$), 평균절대오차(MAE_x)에 대한 가중치(W_x)를 하기의 식들을 통해 산출한다.

$$W_x = 1/(1-R_x^2)$$

[0152] ……..(결정계수 관련 가중치 식)

$$W_x = \log(PVAL_x) * (-1)$$

[0153] ……(유의도 관련 가중치 식)

$$W_x = 1/mae_{x,rev}$$

[0154] ……..(평균절대오차 관련 가중치 식)

[0155] 상기 S850 과정이 완료되면, 상기 가중치 보정부(160)에서 하기의 식을 이용하여 상기 오믹스 데이터의 종류별로 부여된 가중치의 영역별 가중치 평균값(W_{avg})의 지수화를 통해 보정 가중치($w_{x,rev}$)를 산출(S760)한다.

[0156] $$w_{x,rev} = W_{avg}^{(w_x/w_{avg})}$$

[0157] 한편, 상기 가중치 보정부(160)는 각 오믹스 영역의 평균절대오차(MAE_x)에 대한 가중치 보정 전, 실제 나이 분포 대비 평균절대오차를 상대적으로 반영하기 위해 샘플그룹의 평균 나이(AGE_{avg})에 의한 분포보정($mae_{x,rev}$)을 하기의 식을 통해 구할 수 있다.

$$mae_{x,rev} = MAE_x / AGE_{avg}$$

[0158]

[0159] S860 과정이 완료되면, 나이 예측부(170)에서 상기 연관성 분석부의 분석결과 및 상기 오믹스 영역별 데이터에 기초하여 피검자의 나이를 예측하는 구성으로, 하기의 식을 통해 피검자의 나이를 예측(S870)한다.

$$AGE_{integ} = \frac{\sum(AGE_x * w_{x,rev})}{\sum w_{x,rev}}$$

[0160] …….(생체나이 AGE_{integ} 예측식)

[0162] 따라서, 본 발명의 일 실시예를 이용하면, 생체나이 예측모델에 여러 오믹스 영역의 마커를 조합하여 반영함으로써, 개별적인 오믹스 영역에 존재하는 오차를 상쇄하고 더 정확한 생체나이예측이 가능하며, 통합적으로 예측된 생체나이(피검자의 현재 노화상태)에 대하여 각 오믹스 영역의 영향력(또는 노화상태)을 구분하여 해석할 수 있다.

[0163] 예를 들어, 사람의 혈액 등의 샘플의 유전체(텔로미어 길이), 외유전체(메틸레이션), 전사체(유전자 발현) 세 가지 오믹스 데이터 등의 조합을 통해 1. 각 오믹스 데이터에 내재된 노이즈를 상쇄하여 나이예측 정확도를 높일 수 있으며, 2. 피검자의 생체 나이(노화 정도)를 오믹스 별로 구분하여 해석할 수 있다.

[0164] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0165] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

[0166] 100: 시스템

110: 검사샘플 수집부

120: 검사샘플 분석부

130: 전처리 수행부

140: 연관성 분석부

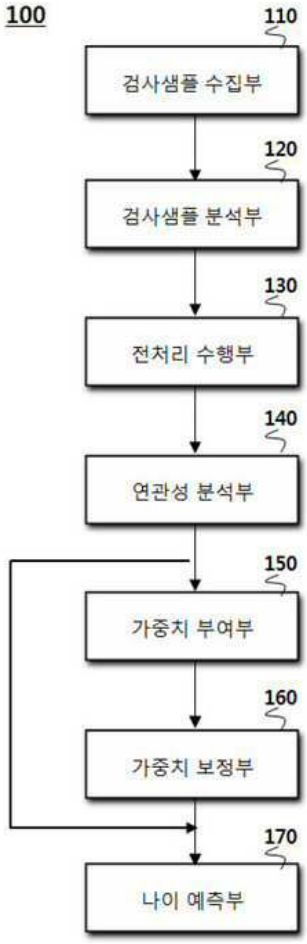
150: 가중치 부여부

160: 가중치 보정부

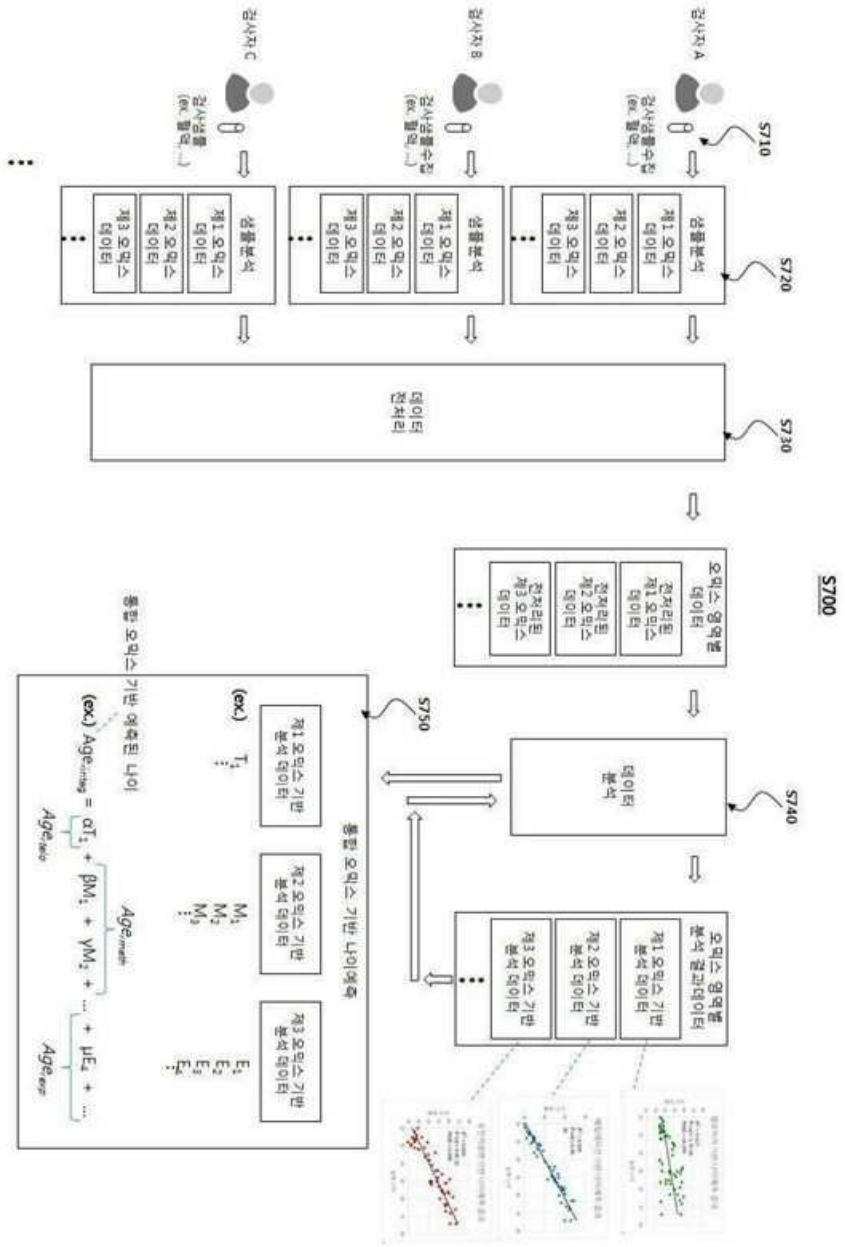
170: 나이 예측부

도면

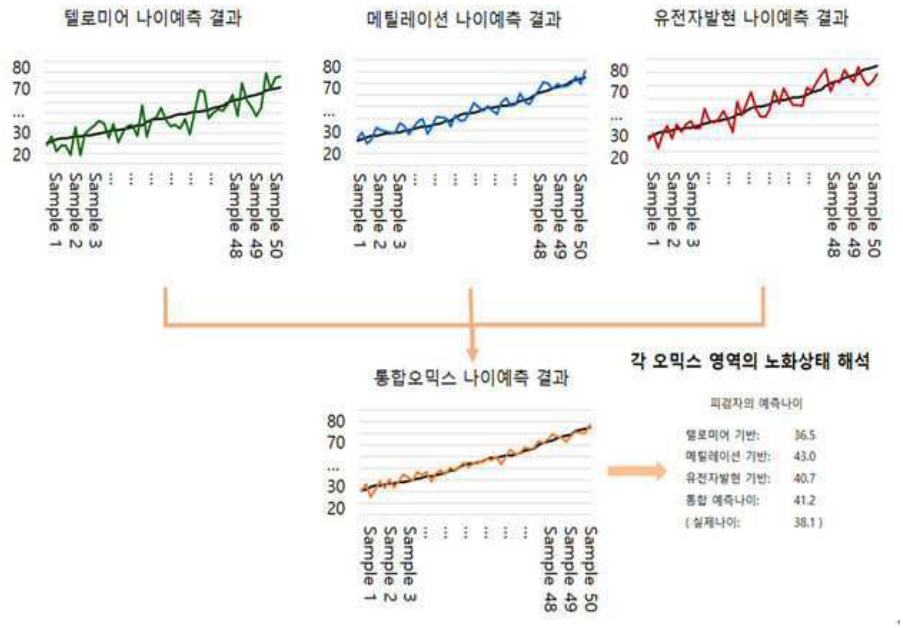
도면1


100
110
검사샘플 수집부
120
검사샘플 분석부
130
전처리 수행부
140
연관성 분석부
150
가중치 부여부
160
가중치 보정부
170
나이 예측부

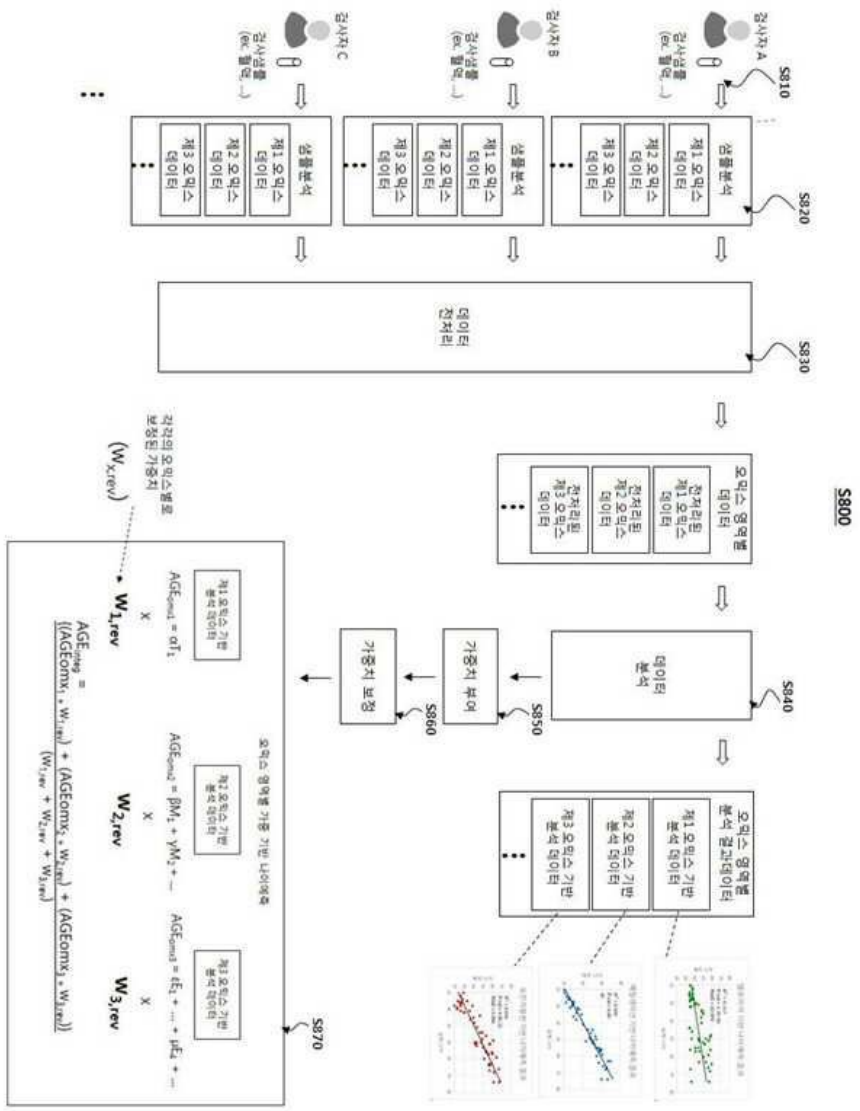
도면2a


S700
S710
S720
S730
S740
S750
검사자 A
검사자 B
검사자 C
검사샘플수집
(ex. 혈액, ...)
샘플분석
제1 오믹스 데이터
제2 오믹스 데이터
제3 오믹스 데이터
데이터 전처리
오믹스 융복합 데이터
전처리된 제1 오믹스 데이터
전처리된 제2 오믹스 데이터
전처리된 제3 오믹스 데이터
데이터 분석
오믹스 융복합 분석 결과데이터
제1 오믹스 기반 분석 데이터
제2 오믹스 기반 분석 데이터
제3 오믹스 기반 분석 데이터
융합 오믹스 기반 나이예측
융합 오믹스 기반 예측된 나이
(ex.)
(ex.) $Age_{ensemble} = \alpha T_1 + \beta M_1 + \gamma M_2 + \ldots + \mu E_4 + \ldots$
Age_{bio}
Age_{meth}
Age_{exp}

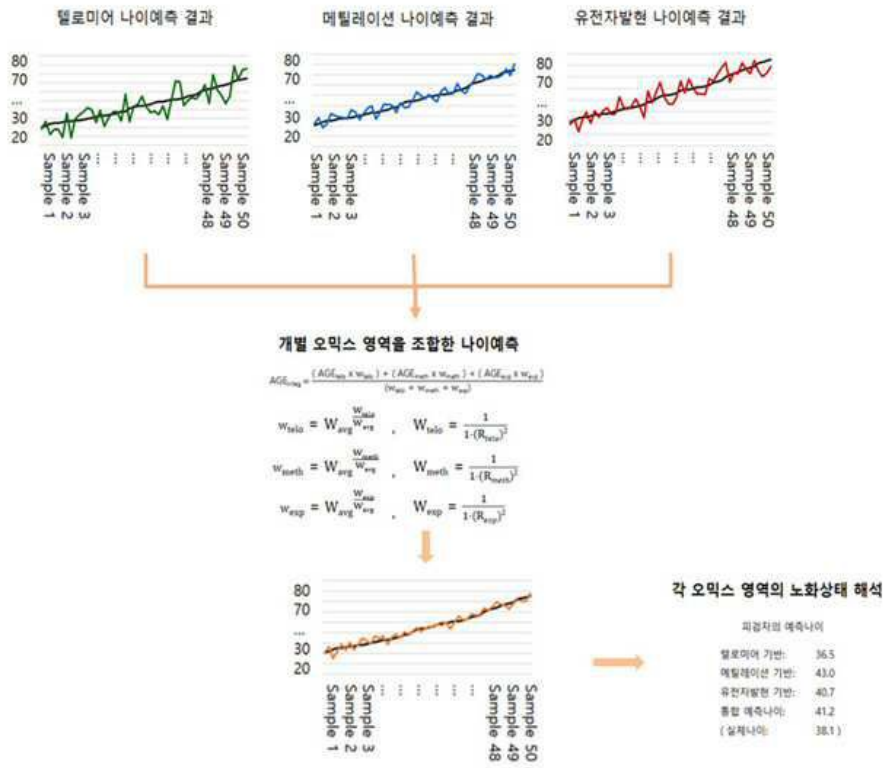
도면2b


텔로미어 나이예측 결과
메틸레이션 나이예측 결과
유전자발현 나이예측 결과
통합오믹스 나이예측 결과
각 오믹스 영역의 노화상태 해석
피검자의 예측나이
텔로미어 기반: 36.5
메틸레이션 기반: 43.0
유전자발현 기반: 40.7
통합 예측나이: 41.2
(실제나이: 38.1)
80
70
...
30
20
Sample 1
Sample 2
Sample 3
Sample 48
Sample 49
Sample 50

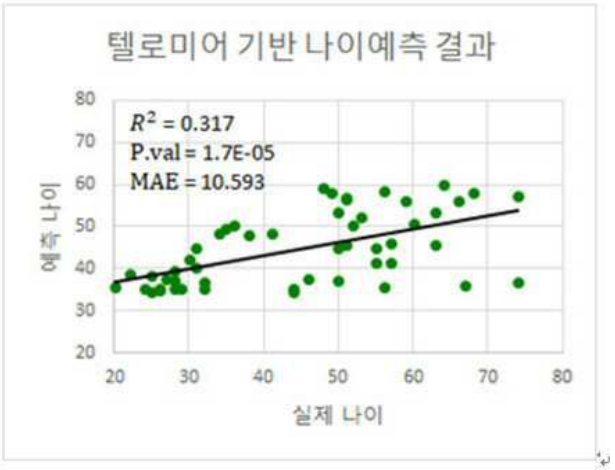
도면2c


S800
S810
S820
S830
S840
S850
S860
S870
검사자 A
검사자 B
검사자 C
샘플분석
제1 오믹스 데이터
제2 오믹스 데이터
제3 오믹스 데이터
데이터 전처리
오믹스 영역별 데이터
전처리된 제1 오믹스 데이터
전처리된 제2 오믹스 데이터
전처리된 제3 오믹스 데이터
데이터 분석
가중치 부여
가중치 보정
오믹스 영역별 분석 결과데이터
제1 오믹스 기반 분석 데이터
제2 오믹스 기반 분석 데이터
제3 오믹스 기반 분석 데이터
각각의 오믹스별로 보정된 가중치
오믹스 영역별 가중 기반 나이예측

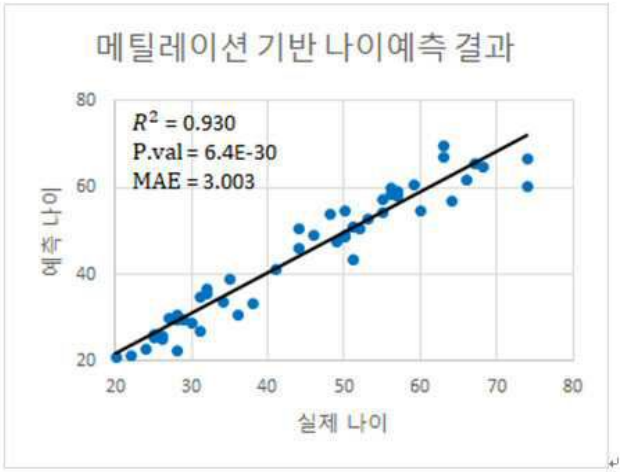
도면2d


텔로미어 나이예측 결과
메틸레이션 나이예측 결과
유전자발현 나이예측 결과
개별 오믹스 영역을 조합한 나이예측
각 오믹스 영역의 노화상태 해석

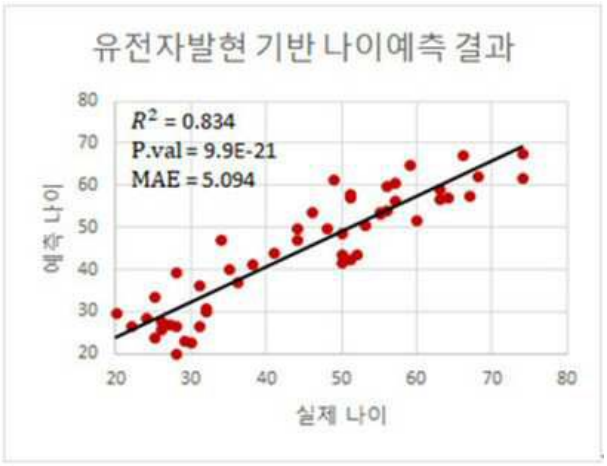
도면3


텔로미어 기반 나이예측 결과
R^2 = 0.317
P.val = 1.7E-05
MAE = 10.593
예측 나이
실제 나이

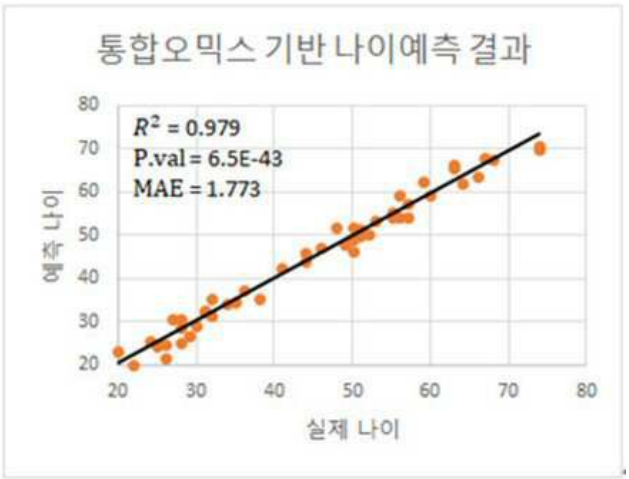
도면4


메틸레이션 기반 나이예측 결과
R^2 = 0.930
P.val = 6.4E-30
MAE = 3.003
예측 나이
실제 나이
20
30
40
50
60
70
80

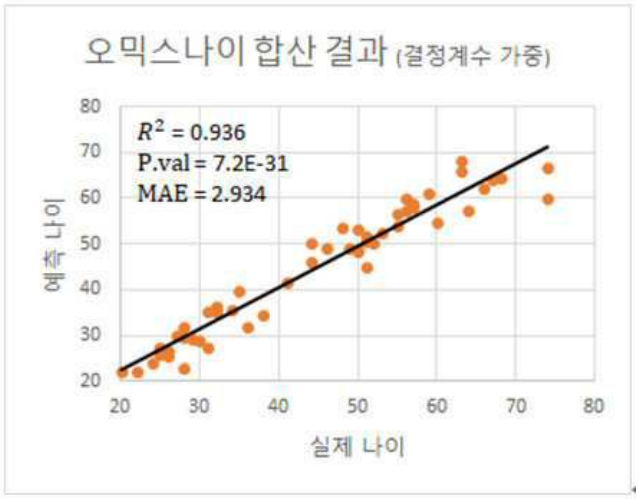
도면5


유전자발현 기반 나이예측 결과
R^2 = 0.834
P.val = 9.9E-21
MAE = 5.094
예측 나이
실제 나이
20
30
40
50
60
70
80

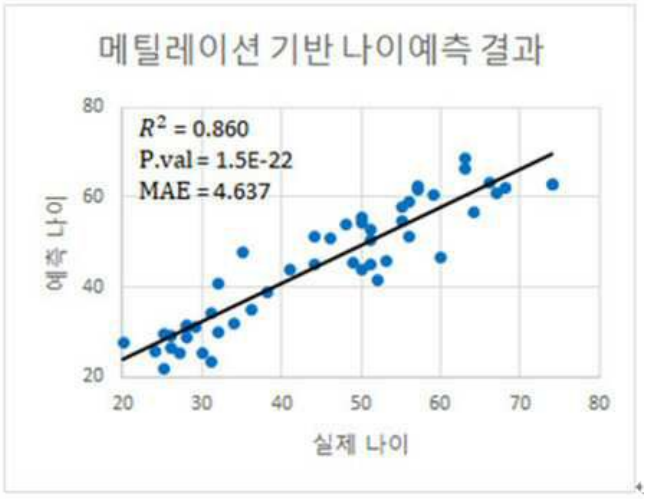
도면6


통합오믹스 기반 나이예측 결과
R^2 = 0.979
P.val = 6.5E-43
MAE = 1.773
예측 나이
실제 나이
20
30
40
50
60
70
80

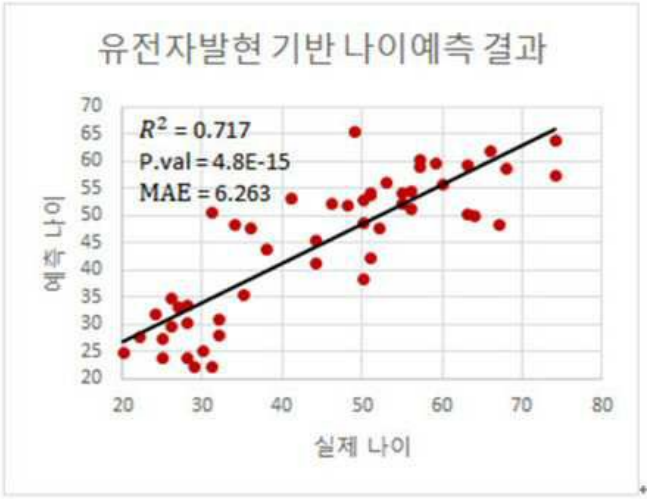
도면7


오믹스나이 합산 결과 (결정계수 가중)
R^2 = 0.936
P.val = 7.2E-31
MAE = 2.934
예측 나이
실제 나이
20
30
40
50
60
70
80

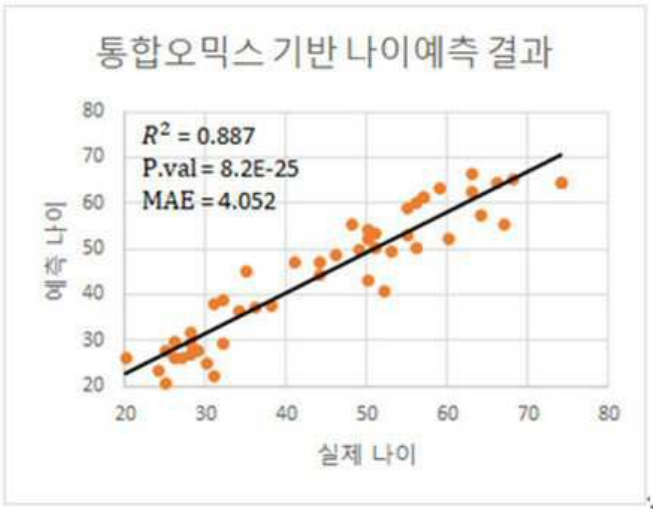
도면8


메틸레이션 기반 나이예측 결과
R^2 = 0.860
P.val = 1.5E-22
MAE = 4.637
예측 나이
실제 나이
20
30
40
50
60
70
80

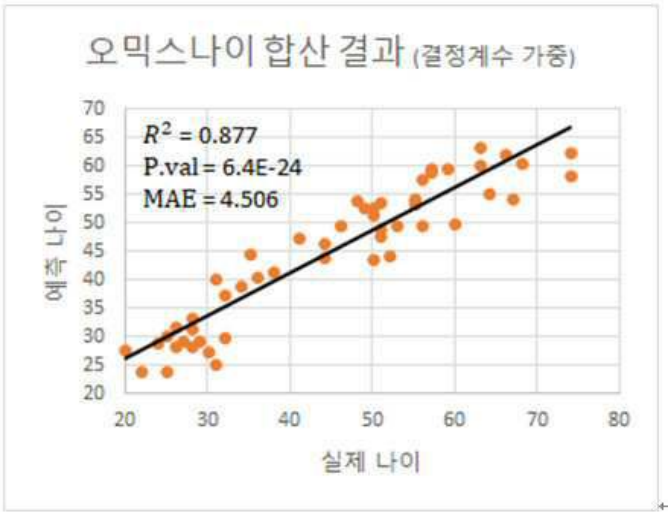
도면9


유전자발현 기반 나이예측 결과
R^2 = 0.717
P.val = 4.8E-15
MAE = 6.263
예측 나이
실제 나이
20
25
30
35
40
45
50
55
60
65
70
80

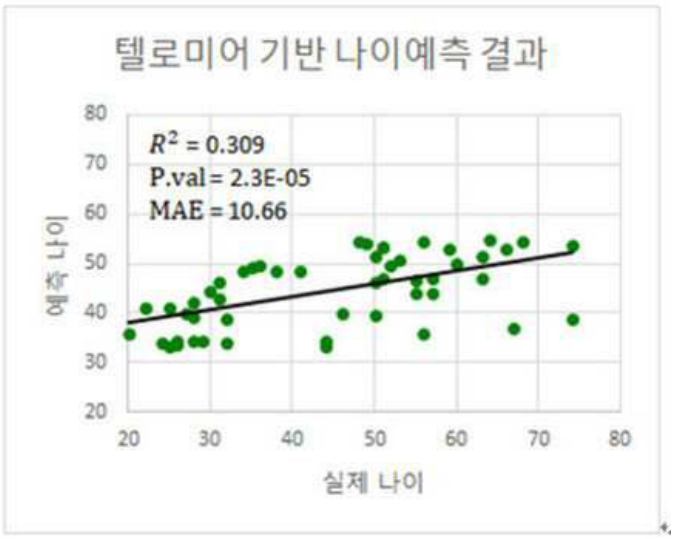
도면10



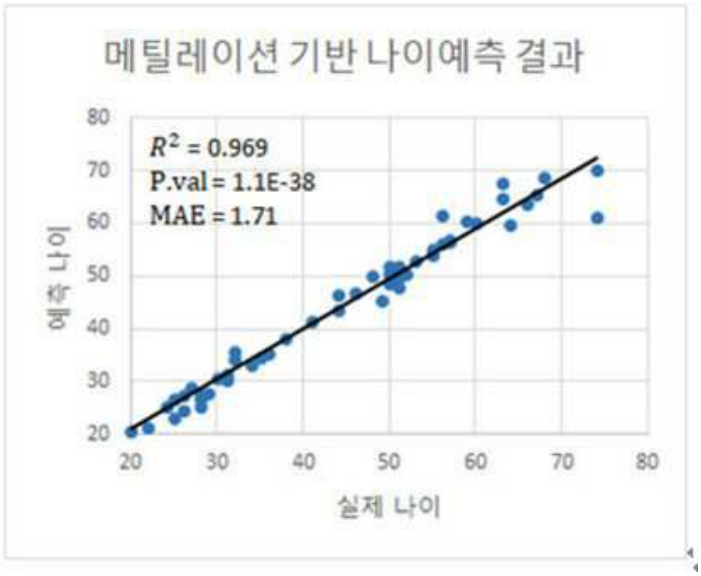
도면11



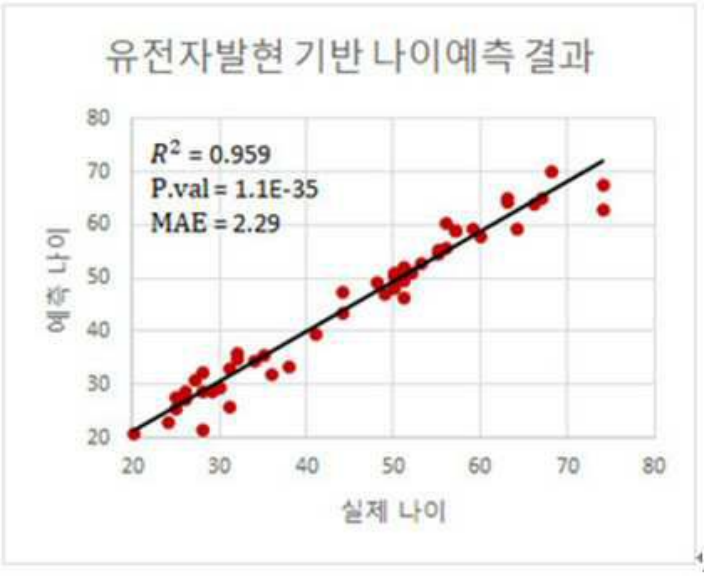
도면12



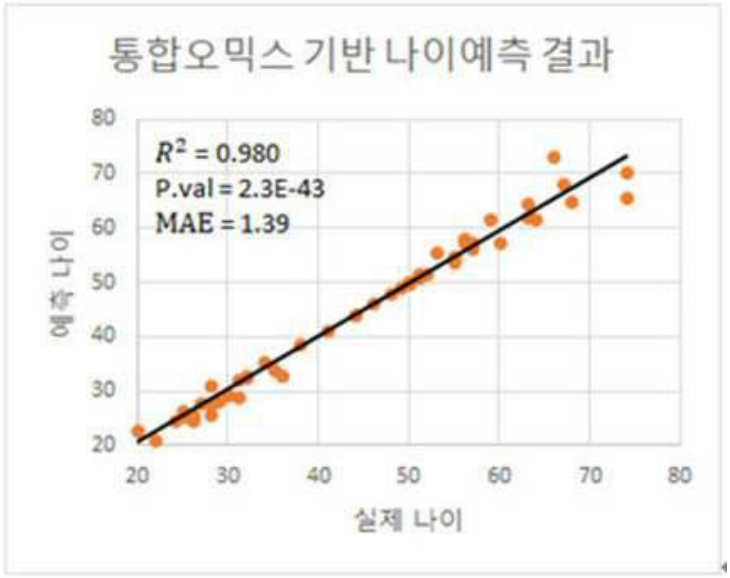
도면13


메틸레이션 기반 나이예측 결과
R^2 = 0.969
P.val = 1.1E-38
MAE = 1.71
예측 나이
실제 나이
20
30
40
50
60
70
80

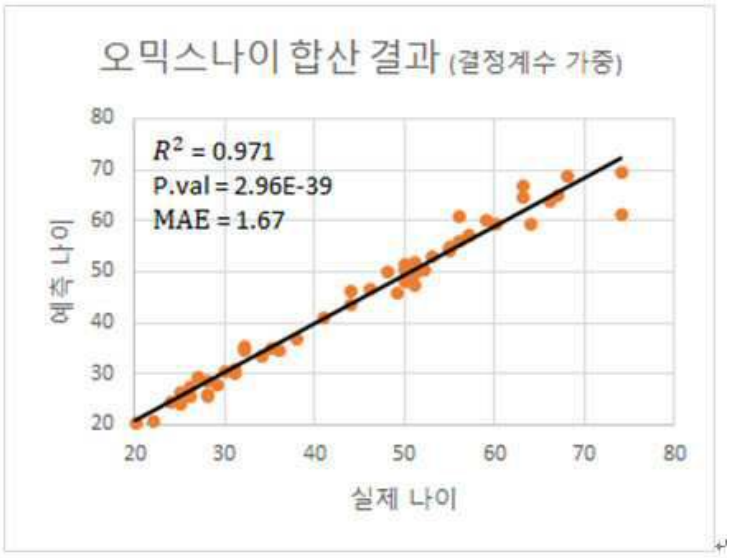
도면14


유전자발현 기반 나이예측 결과
R^2 = 0.959
P.val = 1.1E-35
MAE = 2.29
예측 나이
실제 나이
20
30
40
50
60
70
80

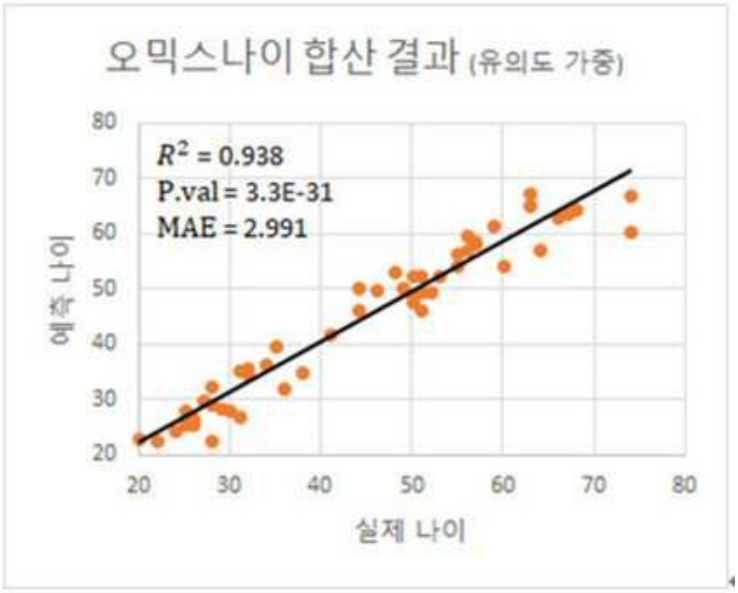
도면15


통합오믹스 기반 나이예측 결과
R^2 = 0.980
P.val = 2.3E-43
MAE = 1.39
예측 나이
실제 나이
20
30
40
50
60
70
80

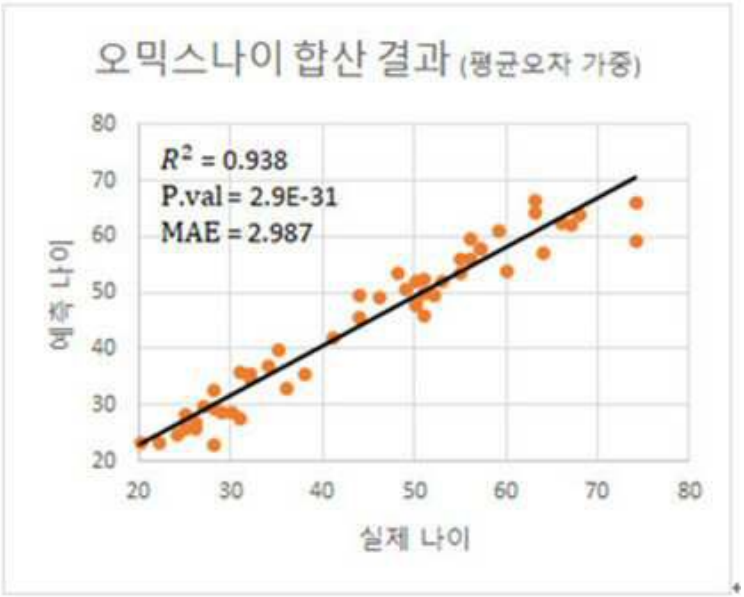
도면16


오믹스나이 합산 결과 (결정계수 가중)
R^2 = 0.971
P.val = 2.96E-39
MAE = 1.67
예측 나이
실제 나이
20
30
40
50
60
70
80

도면17


오믹스나이 합산 결과 (유의도 가중)
R^2 = 0.938
P.val = 3.3E-31
MAE = 2.991
예측 나이
실제 나이
20
30
40
50
60
70
80

도면18


오믹스나이 합산 결과 (평균오차 가중)
R^2 = 0.938
P.val = 2.9E-31
MAE = 2.987
예측 나이
실제 나이
20
30
40
50
60
70
80