

유전자 앱을 만들 수 있는 유전자 앱 센터

권 창 혁^{1,2*}, 함 인 철², 백 석 철², 안 재 균^{1**}

DNA App Center for Creating DNA Apps

ChangHyuk Kwon^{1,2*}, Incheol Ham², Suk-Chul Back² and Jaegyo Ahn^{1**}
netbuyer@inu.ac.kr* and jgahn@inu.ac.kr**

요약

IT 분야에서는 몇몇 앱 스토어나 개발자 센터를 이용하여 많은 앱이 개발되고 있지만, 생물학 분야에서는 2016년 MIT 10대 Breakthrough 기술로 유전자 앱 스토어가 선정되었지만, 생물학자들의 다양한 요구를 충족시키면서 편리한 개발툴이 개발되지 않았고, 아직 초기 단계에 머물러 있다.

우리는 생물학자나 의료 지식을 가진 사람뿐만 아니라 일반인도 쉽게 앱을 만들어서 배포할 수 있는 앱 개발 센터와 유전체 파일을 가진 사람들이 자유롭게 파일을 올려서 다양한 앱을 돌려보고 해석할 수 있는 웹 기반의 DNA App Center(DNA App Center for Developers)를 배포한다. 기본 템플릿을 이용하여 1~2 시간 만에 앱을 제작할 수 있고, 아이디 어반 있으면 프로그래밍을 하지 않고도 디자인, 제작 및 배포까지 가능하다.

ABSTRACT

In the IT field, many apps are being developed using some app stores and developer centers. In the biology field, the DNA App Store was selected by MIT's 10 breakthrough technologies in 2016, but the development tools that meet the diverse needs of biologists, It has not been developed yet and is still in its early stages.

We release a web-based DNA App Center for Developers that allows people with genome files to freely upload files and view and interpret various apps as well as App Builder that make it easy for people to create and distribute apps as well as biologists and medical professionals. You can build your app in 1-2 hours using the default template, and you can design, build and distribute without programming with the idea.

Keywords: App store, DNA App Center, App Builder, SNP marker, DNA App Development Center

I. 서 론

IT(Information Technology) 개발자들은 SK나 삼성의 앱 스토어와 국외의 애플과 구글의 앱 스토어의 앱 개발 플랫폼들을 이용하여 자유롭게 앱을 개발하고 배포 할 수 있다. 대부분의 앱들은 소스 코드를 작성하고 컴파일하여 배포하는 형태이기 때문에 추가적인 파일이나 환경설정을 할 필요 없이 즉시 다운받아서 이용할 수 있다. 최근에는 2D 이미지 추적과 3D 개체 인식을 지원하는 증강현실(AR, Augmented

Reality) 카트와 가상현실(VR, Virtual Reality) 카트뿐만 아니라 GPU에서 개발 및 실행 할 수 있고 인공지능을 지원하는 머신러닝 카트까지 지원해주기 때문에 개발자들은 아이디어만 있으면 누구나 개발과 배포가 가능하다.

생물학과 의료 분야에서는 유전자와 유전체의 다양한 파일 포맷과 최소 600메가 이상의 VCF(Variant Call Format) 파일의 처리 문제와 정형화 되지 않은 프로그램들로 인해서 기존의 IT 앱 개발자 플랫폼을 이용하기는 어려운 실정이다. 수백억을 쓴 많은 GWAS 연구 결과를 모아 놓은 GWAS

* 1저자, **교신저자

¹ 인천대학교 컴퓨터공학과, ² 마이지놈박스

논문접수일:2018.11.06. 논문심사일:2018.12.01. 게재확정일:2019.01.12.

Central[1], GWAS Catalog[2], GWASdb2[3]등이 있고, 대부분의 사람들이 관심이 없을 수 없는 암, 건강, 헬스케어, 의료 관련된 수많은 의료 콘텐츠 제품들이 존재하지만, 관련 앱을 개발하고 배포 할 수 있는 개발 플랫폼은 거의 없는 실정이어서, GUI환경에서 쉽게 개발, 테스트 및 배포 할 수 있는 개발 플랫폼이 있다면 의/생물학자들에게 매우 유용한 수 있다.

본 논문에서는 앱 개발자들이 GWAS 연구 기반으로 밝혀진 수많은 마커를 이용하여 콘텐츠를 개발하고, 기본적인 디자인 템플릿을 이용하여 앱을 배포 할 수 있을 뿐만 아니라 접속, 구매, 판매에 관한 통계정보를 쉽게 확인할 수 있는 유전자 앱 개발 플랫폼인 DNA App Center(DNA App Center for Developers), <https://developer.mygenomebox.com/>를 개발하여 배포한다. 프로그램 랭귀지를 공부하지 않은 생물학자들도 아이디어만 있으면 2시간 이내에 single/multiple marker를 이용하여 자유롭게 앱을 개발할 수 있는 최초의 플랫폼이다.

II. 관련 연구

IT 앱 공급 업체들은 모바일 앱 생태계의 경쟁이 심화됨에 따라 사용자의 다양한 경험 정보를 얻기 위한 전략으로 App Store 2.0[4]의 개념을 이용하여 앱, 기기 및 사용자에 관한 크리우드소스 정보(Crowdsourced Information)를 활용하여 모바일 앱의 품질을 향상시키고 있으며 군중 정보로부터 다양한 종류의 피드백을 주고받는 방향으로 나아가고 있다. 이런 피드백을 활용하여 개발자가 앱을 제시하기 전에 앱에 영향을 미칠 수 있는 잠재적 오류를 줄이고 사용자의 앱 다운로드에서도 에러를 줄일 수 있다. 다양한 모바일 개발을 지원하기 위해서 머신 러닝 API를 지원함과 동시에 얼굴 인식에 월등한 성능을 보이는 CNN(Convolutional Neural Network)[5], 딥러닝, 가상 현실과 증강현실을 지원하여 화려한 영상과 3차원의 360도 지원으로 몰입도를 최대로 이끌어내는 기능들을 내장하고 있어서 개발자들은 최대한의 능력을 발휘할 수 있다. 또한, 시계와 헬스케어 기능을 접목한 애플 Watch[6]와 삼성 gear는 핸드폰의 기능을 피드백스 드레이와 커뮤니케이션 장치로 확장하면서 건강 관련 미니 앱 스토어 역할을 한다. 업로드와 다운로드에 관한 통계정보들을 활용하기 위해서 생태계[7] 논문에 영감을 얻은 AppEco[8]는 개발자 에이전트(developer agents)와 사용자의 성장을(growth rate)을 시그모이드 성장 함수(sigmoid growth function)를 적용하여 랭킹 알고리즘을 적용하면서 개발자와 사용자의 정보를 활용하여 판매와 마케팅에 사용하여 앱 생태계를 구축하려는 노력도 하고 있다.

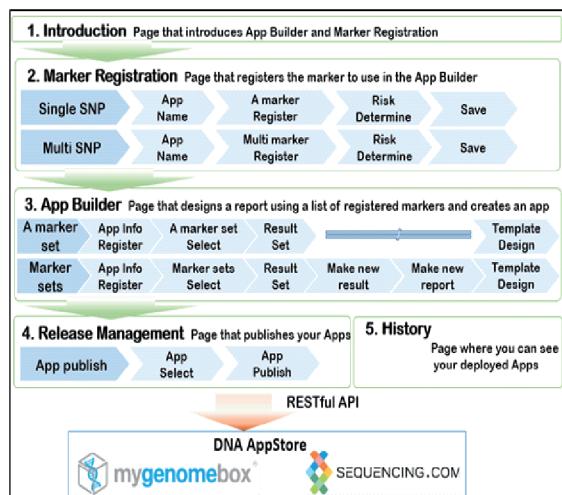
IT 앱 스토어가 국내외 대기업들의 활발한 기업 활동으로 급성장하고 있지만 유전자 앱 스토어는 2016년 MIT 10대 Breakthrough로 지정은 되었지만, 기술에 대한 논문은 거의 보고 되고 있지 않다. 모든 분석을 시작하기 전에 유전체 파일이 필요한데 최소 600메가에서 1200메가의 VCF(Variant Call Format) 파일의 처리 문제와 정형화 되지 않은 프로그램들로 인해서 기존의 IT 앱 개발자 플랫폼을 이용하기는 어려운 실정이다. 미국 벤처 기업인 Sequencing.com과 Helix가 유전자 앱 스토어를 운영하고 있지만, 대부분 API 기반의 앱 마켓에 한정되어 있고, 개발자들이 유전자 앱을 만들 수 있는 공간은 전혀 없다. 국내는 한 제품만을 개발하고 판매하는 형식으로 발전되어 왔다. 개발된 앱을 연결하기 위한 시도는 API를 이용한 2가지 방법으로 시도 되고 있는데, 첫번째 방법은 App Market API인데 서비스 회사는 개발자에게 고객의 파일을 전송하고, 개발자는 결과 레포트를 전송해주는 역할을 하는 딘슨 API이고 개발자가 개발해 놓은 앱을 연결하는 역할만 한다. 두번째 방법으로 +RTP(Real-Time Personalization) API는 날씨와 같은 실시간 정보를 빠르게 받을 수 있도록 사용자와 앱 개발자가 알고자 하는 유전자 정보나 설문 정보를 실시간으로 빠르게 확인할 수 있는 API이다. 실시간 정보를 건강, 영양, 라이프스타일 서비스 업체들에 전달함으로서 유전 정보와 실생활 정보를 결합한 통합 패키지 형태의 서비스를 제공하기 위한 목적이이다.

III. DNA App Center

3.1 시스템 구성

Application 메뉴는 5개의 하위 메뉴로 구성되어 있고, 총 6단계의 Step으로 자신만의 앱을 만들 수 있다. Introduction 메뉴에서는 기본적인 앱 제작 순서가 간단히 설명되고, Marker Registration 메뉴는 웹에서 사용할 수 있는 마커를 등록하는 단계인데, 하나의 마커만을 이용하여 앱을 만들 경우는 Single Marker(Single SNP, Single Type Marker)를, 2개 이상의 마커를 사용하여 앱을 만들 경우는 Multiple Marker(Multi SNP, Multi Marker, Multiple Type Marker)를 이용한다. 유전자 이름(Gene Name)과 SNP 아이디(rsID Number), Normal Allele 타입, Effect Allele(Risk Allele) 타입, Order Ratio 영역을 지정하고 최종적으로 Order Ratio가 가지는 값의 전체 영역을 이용하여 Risk 구간(score)를 구분하여 설정한다. App Builder 메뉴에서는 Marker Registration 단계에서 만들었던 마커 셋과 결과 레포트를 연결하여 모든 앱 제작을 완료하는 단계이다. 총 6단계의 Step이 있는데 이 전에 만들어 놓은 marker나 결과가 없으면 4단계(Sctp4,

Enter mapping data results)와 5단계(Step5, Set overall result page)는 Skip하여 4개의 단계 만에 모든 앱 제작을 완료할 수 있다. Release Management 메뉴에서는 App Builder에서 제작한 앱을 App Center 심사위원에게 승인을 요청하는 단계이고, 심사위원들은 앱에 기술한 내용과 문법, 레포트의 구성과 참조한 논문의 진위 여부 등을 검토하여 앱 개발자에게 승인이나 수정을 요청한다. History 메뉴에서는 제작한 앱 아이디와 승인상태 등을 확인 할 수 있다. 랭귀지는 php, 베이타베이스는 mariaDB를 이용하였으며 DNA App Center, <https://developer.mygenomebox.com/>의 Developer Console에서 Application 메뉴의 모든 기능을 이용 가능하다.



[그림 1] 앱 제작을 위한 절차

[Fig.1] Procedures for building your app

3.2 앱 제작 예시

Single marker와 Multiple marker로 구분하여 등록 및 제작이 가능하다. Single marker는 액취증(Deodorant, Rs17822931)이나 알코올 흥조(Alcohol flush)등과 같이 하나의 마커도 전체를 설명할 수 있는 경우에 Single marker를 이용하여 앱을 제작하고 Multiple marker는 대머리 관련 유전자들, 눈 색깔 관련한 유전자들, 당뇨병과 같이 복합적인 유전자가 작용할 때는 이 기능을 이용하여 앱 제작이 가능하다. <그림 1>에서 2번의 Marker Registration 메뉴에서 Make Markers를 클릭하면 마커 셋에 관한 이름과 설명을 기술하는 페이지가 나오고 SAVE&NEXT를 클릭하면 다음 Single인지 Multiple인지를 선택하는 봄보 박스가 나오고 용도에 맞게 기술 할 수 있다.

3.2.1. Single Marker

<그림 2>의 Single marker는 하나의 rsID와 유전자만 기술하며 두 개의 allele 리스트는 연구를 참고하여 Normal과 Risk allele을 기술한다. Order ratio(OR) 값은 NN은 2개 모두 Normal allele를 가진 경우의 위험도를, NR은 하나의 Normal과 하나의 Risk allele를 가진 경우의 위험도를, RR은 두 개 모두 Risk allele를 가진 타입일 때의 위험도를 기술한다. 위험도의 score는 논문에 기술된 order ratio를 그대로 기술할 수도 있고 p-value를 참고하여 적절한 구간을 선정하여 나눌 수도 있으며 SNPedia 사이트나 논문들을 참고하여 알고리즘을 구현하여 구간을 설정할 수도 있다. 위에서 기술한 위험도 값이 Value에 그대로 들어가지만 수정 가능하고 사용자에게 보여질 결과는 Description에 기술하면 되는데 사용자는 이런 결과를 신뢰하기 때문에 심사숙고하여 Value, Risk Name, Description을 기술하여야 한다.

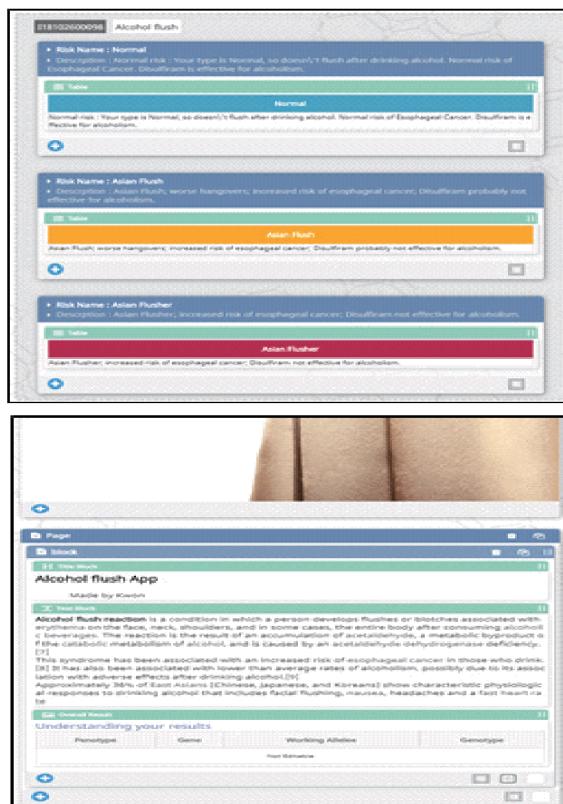
| Item Name | Alcohol flush | | | | | | |
|--------------|---|--|---------------|-------------|-------|-------|-------|
| Item Type | <input checked="" type="radio"/> Single SNP <input type="radio"/> Multi SNP | | | | | | |
| Item Version | 1.0.0 | | | | | | |
| Gene | rs671 | rsID | Normal Allele | Risk Allele | OR(N) | OR(M) | OR(R) |
| ALDH2 | rs671 | G | A | 1 | 3.5 | 4 | |
| Value | Risk Name | Description | | | | | |
| 1 | Normal | Normal risk : Your type is Normal, so doesn't flush after drinking alcohol. Normal risk of Esophageal Cancer. Disulfiram is effective for alcohol flush. | | | | | |
| 3.5 | Asian Flush | Asian Flush; worse hangovers; increased risk of esophageal cancer; Disulfiram probably not effective for alcoholism. | | | | | |
| 4 | Asian Flusher | Asian Flusher; increased risk of esophageal cancer; Disulfiram not effective for alcoholism. | | | | | |

[그림 2] 싱글 마커의 기본적인 내용

[Fig.2] Basic contents of a single marker

<그림 1>의 3번의 App Builder는 4단계로 구성되어 있으며, <그림 2>에서 제작한 마커와 위험도를 연결하며 레포트의 최종 디자인을 하는 단계이다. 1단계는 App Name, App Description, Gender, Category, Nation, Race Group, App Price, Icon등의 14가지를 입력하는 단계이다. Gender는 여성만 가지는 특징인 조기폐경이나 유방 크기와 같은 앱을 제작할 경우 반드시 Female을 체크하고 남성, 여성 모두 가능하면 All을 체크한다. Nation은 서비스를 하고 싶은 나라를 체크하며 법률이나 규제 등의 이유로 특정 나라를 제외할 수도 있다. Race Group은 연구가 특정 인종에 의존적일 때 체크를 하며 몇몇 표현형은 인종에 상당히 의존적이기 때문에 필수로 체크를 하는 것이 필요하다. 앱 가격은 Free App을 지향하며 많은 노력이 들어간 경우에는 가격을 개발자가 결정할 수 있고 App Version은 초기에는 반드시 1.0.0을 기술하여 다음단계로 넘어갈 수 있다. 2단계는 <그림 2>에서 제작한 마커를 선택하면 된다. 3단계인 <그림 3>의 결과 설정(위)은

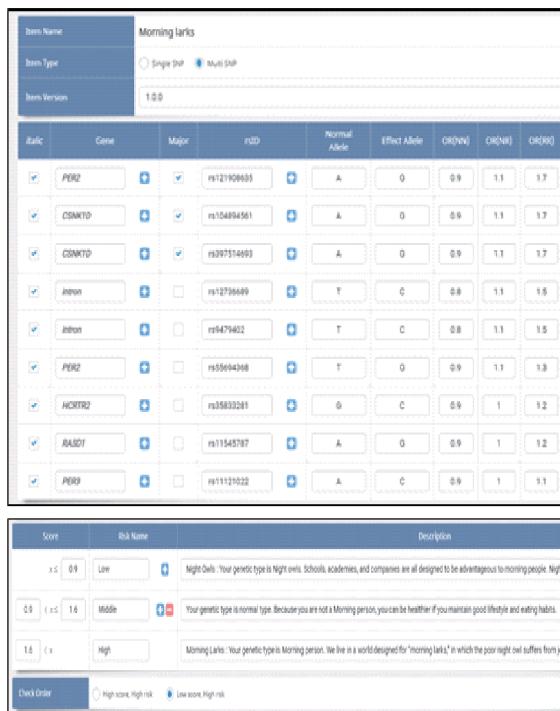
마커에서 기술한 Description이 그대로 나오는데 Risk Name이 보이는 컴포넌트의 펜 기능을 이용하여 최종 결과를 수정할 수 있다. <그림 3>의 위쪽인 템플릿 디자인 페이지는 마지막 4단계이며 DNA App Reporting Builder에서 제공된 템플릿을 이용할 경우는 Add Template를 클릭하고, Cover와 Page의 모든 내용을 스스로 디자인 한 경우에는 Add Element를 클릭하여 디자인과 내용에 맞게 삽입할 수 있다. Cover는 앱의 첫 페이지이기 때문에 앱의 이름과 앱의 요약한 내용과 배경에 적합한 이미지를 삽입할 수 있고, Page는 앱에 대한 전체적인 설명과 연구 내용에 대한 내용이 들어가며 Block 컴포넌트위에 Title Block(앱의 주제), Text Block(연구 내용), Image Block(비교 결과나 적절한 이미지)을 추가하여 자유롭게 배치 순서와 모양을 변경할 수 있다. 3단계에서 설정한 위험도에 대한 결과 판단은 Overall Result로 제일 아래쪽에 자동으로 첨부된다.



[그림 3] 앱 빌더의 결과 세팅(위)과 디자인 템플릿(아래)
[Fig.3] App Builder settings(Up), design template(Down)

3.2.2. Multiple Markers

하나의 중요 마커 보다는 여러 개의 마커로 표현형이나 질병을 해석해야 할 경우에 <그림 4>의 Multiple maker 기능을 이용한다. 여러 개의 마커를 입력할 수 있으며 반드시 체크해야 될 마커는 Major를 체크해야 하며, 사용자의 유전자 파일에 마커가 존재 하지 않을 경우에는 분석을 할 수 없으므로 주의해서 체크해야 한다. 여러 개의 마커가 가지는 값의 범위가 다양하므로 최소값과 최대값의 범위를 체크하여 3단계(Low, Middle, High)로 나눌 수도 있고, 5단계의 복잡한 단계로도 구분할 수 있다. 마지막의 Check Order는 위험도가 높을 때 High로 표기하고 High score(High risk)를 체크하며 Low score일 때 High risk를 표기해야 할 경우는 Low score를 체크한다. Builder는 <그림 5>에서 제작한 마커 셋과 위험도를 연결하며 레포트의 최종 디자인을 하는 단계인데, <그림 4>의 하나의 블러플 마커 셋 만을 이용할 경우는 총 4개의 단계 만에 제작할 수 있지만, 2개의 마커 셋을 합할 경우는 6단계로 작업해야 한다. 한 개의 마커 셋 만을 이용할 경우는 1단계에서 마지막 단계 까지는 Single marker와 동일하다. 2개의 마커 셋을 이용할 경우 결과 매칭단계(Enter mapping data results)를 포함하여 기준과 유사하게 3단계(Normal, Middle, High)로 위험도를 표기하고 Set Result Values를 이용하여 조합된 결과를 최종 3단계로 설정한 Risk type에 재 매핑을 해주어야 한다. 새로운 논문이나 결과가 추가되었을 때 주어진 마커 만 삽입하여 레포트를 유지할 수 있는 장점이 있다. Set overall result page 단계에서 결과 용어나 내용을 수정하여 마지막 단계인 템플릿 디자인 단계에서 모든 디자인을 마무리 할 수 있다. 마지막 단계는 DNA App Reporting Builder에서 제공하는 템플릿(Add Template)이나 Add Element를 이용하여 모든 디자인과 결과를 매핑하여 최종 레포트를 완성 할 수 있다.



[그림 4] 멀티 마커의 기본적인 내용
[Fig.4] Basic contents of multi-marker

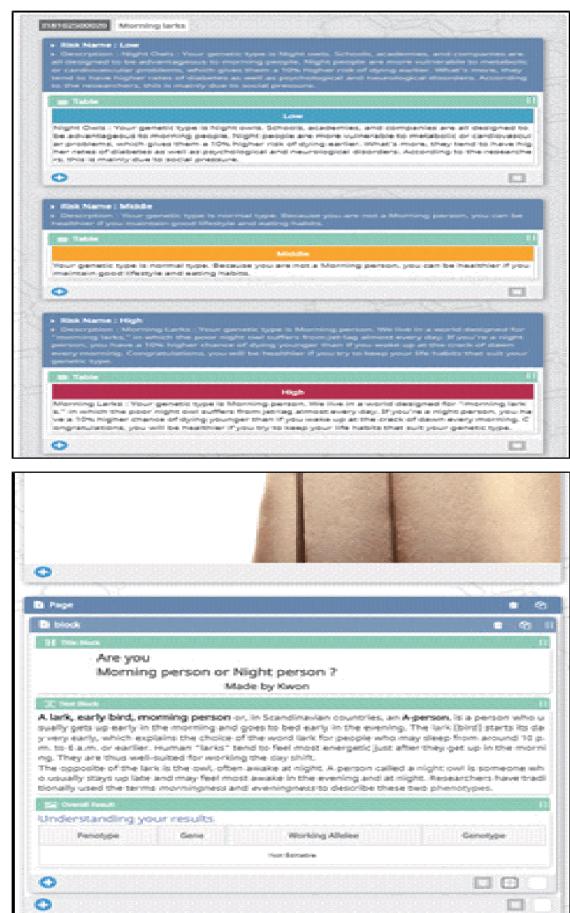
IV. 결론 및 향후 연구

DNA App Center를 이용하여 Single marker를 이용하여 앱을 만드는 예시는 Alcohol flush를 구현하였고, Multiple marker 예시는 Morning larks(person)을 구현해 보았다. 구성한 앱을 가장 잘 표현하는 디자인을 직접 구성하여 아름다운 앱을 구현할 수도 있지만, 제공해주는 기본 템플릿만을 이용하여도 충분히 앱 제작이 가능하다. 처음 메뉴얼 공부나 가능 익히는 연습만 한다면 1~2 시간 반에 앱을 디자인, 제작 및 배포까지 충분히 가능하고 아이디어만 있으면 IT 앱 같이 직접 프로그래밍을 하지 않아도 된다는 장점이 있다.

GWAS 연구의 결과인 GWAS Central[1], GWAS Catalog[2], GWASdb2[3]등에 등록되어 있는 수많은 마커들로 자유로운 앱을 구현 가능하고, 처음에는 한 두개의 SNP 마커만을 이용하여 배포하다가 향후에는 Multiple marker로 복잡한 앱 개발도 가능하다. 시рин싱 가격이 떨어지고 대부분의 사람들이 유전자 파일을 가지고 있을 몇십년 이내에는 전문 직업으로도 가능할 것으로 예상한다.

생물학자나 연구자들이 올리는 간단한 앱도 중요하지만,

향후에는 Multiple marker들이 상호 복합적으로 작용하는 관계를 반영할 수 있는 알고리즘과 생물정보학자나 전문 프로그래머들이 개발하여 유전체 전체를 해석할 수 있는 머신러닝이나 딥러닝이 적용된 알고리즘 앱도 개발되어서 보다 정확한 해석이 필요할 것이다. 주제별로 제공되는 많은 템플릿이 구성되어 있어서 아이디어만 있으면 디자인 걱정 없이 자유롭게 앱을 올릴 수 있는 환경도 더 많은 생물학자들과 좋은 앱 개발자들을 모으는 중요한 요소가 될 것이다.



[그림5] 앱 빌더의 결과 세팅(위)과 디자인 템플릿(아래)
[Fig.5] App Builder settings(Up), design template(Down)

참 고 문 헌

- [1] Tim Beck, Robert K Hastings, Sirisha Gollapudi, Robert C Free, and Anthony J Brookes, "GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies," *Eur J Hum Genet.* 22(7): 949 - 952. July 2014.
- [2] MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington ZM, Welter D, Burdett T, Hindorff L, Fllicek P, Cunningham F, Parkinson H, "The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)," *Nucleic Acids Res.* 4:45(D1):D896-D901. January 2017.
- [3] Mulin Jun Li, Zipeng Liu, Panwen Wang, Maria P. Wong, Matthew R. Nelson, Jean-Pierre A. Kocher, Meredith Yeager, Pak Chung Sham, Stephen J. Chanock, Zhengyuan Xia, and Junwen Wang, "GWASdb v2: an update database for human genetic variants identified by genome-wide association studies," *Database* 44(Database issue): D869 - D876. January 2016.
- [4] María Gómez, Bram Adams, Walid Maalej, Martin Monperrus, Romain Rouvoy, "App Store 2.0: From Crowdsourced Information to Actionable Feedback in Mobile Ecosystems," *IEEE Software* Volume: 34, Issue: 2 Page s: 81-89 Mar.-Apr. 2017.
- [5] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, 86(11):2278-2324, November 1998.
- [6] Katina Michael and M.G. Michael "Apple Watch Temptation," *IEEE Consumer Electronics Magazine* Volume: 4, Issue: 4, October 2015.
- [7] S. E. Kingsland, "Modeling Nature: Episodes in the History of Population Ecology," University of Chicago Press, 1995.
- [8] Lim, S. L., & Bentley, P. J. Lim, S. L., & Bentley, P. J., "Investigating app store ranking algorithms using a simulation of mobile app ecosystems," *2013 IEEE Congress on Evolutionary Computation* Page s: 2672 - 2679 2013.



권창혁 (ChangHyuk Kwon)

2014년~2016년 Eone-Diagnostics
Genome Center

2014년 삼성유전체 연구소

현재 MyGenomeBox

※ 관심분야 : 생물정보학, 인공지능

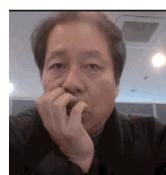


함인철 (Incheol Ham)

2016년~현재
マイジ늄박스

현재 MyGenomeBox

※ 관심분야 : 생물정보학, 전산개발, 비트코인



백석철 (Suk-Chul Back)

2012년~2017년 엔아이비디랩 대표이사
2010년~2011년 (주) SK STMS 개발팀장
2008~2010년 (주) NHN 소셜네트워크

현재 MyGenomeBox

※ 관심분야 : 빅데이터, 블록체인



안재균 (Jaegyo Ahn)

2006년 연세대학교 컴퓨터공학과(공학사)

2009년 연세대학교 (공학석사)

2013년 연세대학교 (공학박사)

현재 인천대학교 컴퓨터공학과 교수

※ 관심분야 : 생물정보학, 데이터마이닝, 기계학습, 전산생물학