



## 저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#) 

강창수 교수 지도

석사학위 청구논문

차세대 염기서열 분석기술을 활용한  
Amyloidosis Cutis Dyschromica의  
원인 유전자 규명

2016

성신여자대학교 대학원

생물학과

김 지 나

차세대 염기서열 분석기술을 활용한  
Amyloidosis Cutis Dyschromica의  
원인 유전자 규명

강창수 교수 지도

이 논문을 석사학위논문으로 제출함

2015년 11월

성신여자대학교 대학원

생물학과

김 지 나

# 인준서

김지나의 석사학위 논문으로 인준함

2015년 11월

심사위원장 \_\_\_\_\_ (인)

심 사 위 원 \_\_\_\_\_ (인)

심 사 위 원 \_\_\_\_\_ (인)

성신여자대학교 대학원

## 논문개요

Amyloidosis Cutis Dyschromica(ACD)는 피부색소가 불균등하게 분포됨으로써 정상 피부색보다 어두운 색 및 밝은 색 반점이 피부 전반에 나타나는 질환이다. 지금까지 전세계적으로 약 48여명의 환자들이 보고되었으며, 대부분의 환자들이 가족력을 가진 것을 고려할 때 유전질환인 것으로 추측되고 있지만, 지금까지 원인 유전자는 알려져 있지 않았다.

본 연구에서는 ACD의 원인유전자를 규명하는 것을 목표로 설정하였으며, 이를 위해 파키스탄에서 3명의 ACD환자가 발생한 동계교배 가계도를 수집하였다. 파키스탄에서는 전통적으로 동계결혼이 성행하고 있는데, 이로 인해 가족 내에서 열성유전병이 빈번하게 나타날 수 있으므로, 유전병 연구에 있어 유용한 점이 있다. 수집된 가계도내에서 유전양상을 살펴보았을 때 ACD는 열성 유전되고 있는 것으로 가정하고 연구를 수행하였다.

가계도 구성원들의 genomic DNA를 대상으로 염기서열 정보를 분석하기 위해 차세대 염기서열 분석기술 (Next-Generation Sequencing, NGS)의 일종인 엑솜시퀀싱 (Whole-Exome Sequencing, WES)을 수행하였다. Bioinformatic 단계에서는 DNASTAR사의 Lasergene 12 core suite 프로그램 패키지를 활용하여 ACD 환자들의 엑솜 염기서열을 분석 하였으며, 가계 구성원들로부터 총 2,369,697개의 염기서열 변이들을 발견하였다. 가계도내에서 ACD의 유전 패턴, 염기서열 변이들의 빈도 및 *in-silico* 분석결과 *Glycoprotein non-metastatic melanoma protein b* (GPNMB) 유전자에서 발견된 c.1124G>T 돌연변이가 ACD발병의 유전적 원인인자인 것으로 나타났다. 이 돌연변이는 GPNMB 단백질의 375번째 아미노산인 글리신을 발린으로 치환시키는 missense 돌연변이 (GPNMB p.Gly375Val)인데, 염기서열변

이 DB(1000genome 및 ExAC)에 지금까지 보고되지 않은 희귀 돌연변이인 것으로 나타났다.

지금까지 인간에서 *GPNUMB* 유전자의 기능 이상이 어떤 질환을 초래하는지는 전혀 알려져 있지 않았지만, 피부색소인 melanin의 형성 과정에 관여하고 있음은 이미 잘 알려져 있다. 따라서, *GPNUMB*의 기능변이로 인해 ACD 질환이 발병할 가능성은 매우 높다. 그러나, 보다 구체적인 *GPNUMB*의 작용 기작은 잘 알려져 있지 않아, 후속 연구가 진행될 필요성이 있다.

# 목 차

논문개요

목차

그림 목차

표 목차

I. 서론	1
II. 이론적 배경	
1. Amyloidosis Cutis Dyschromica	5
2. 멘델의 열성유전 모델	7
3. 유전학연구 측면에서 동계결혼의 이점	8
4. 돌연변이의 종류	9
5. 생물정보학을 이용한 염기서열 분석	
(1) 생물정보학(Bioinformatics)	10
(2) 차세대 염기서열 분석 (Next-Generation Sequencing)	12
(3) 엑솜 염기서열 분석법 (Whole-exome sequencing)	13
6. 희귀유전질환	15
III. 연구방법	
1. 연구대상	16
2. 가설	18
3. 연구방법	
(1) DNA 정제 및 엑솜 캡처링	18

(2) DNASTAR Lasergene 12 core suite program .....	19
(3) 유전정보 검색에 유용한 데이터베이스 .....	20
(4) MAF(Minor allele frequency) 조사 .....	22
(5) 파키스탄 인종 내 유전자형 분석 .....	23
(6) 돌연변이로 인한 단백질 기능 변화 예측 .....	25
(7) Vector 디자인 .....	26

#### IV. 결과

1. ACD 환자의 임상적, 조직학적 특징 .....	28
2. 가계도 내 염기서열 분석 결과 .....	34
3. 파키스탄 인종 내 유전자형 분석 .....	39
4. 돌연변이로 인한 단백질 기능 변화 예측 .....	40
5. 플라스미드 내 돌연변이생성(Mutagenesis) .....	41
6. 유전자 발현 실험 .....	42

#### V. 결론 및 논의 .....

(1) <i>GPNMB</i> 유전자 .....	49
(2) 멜라닌 소체 및 멜라닌 형성 .....	52

#### 참고문헌

#### ABSTRACT

## 그림 목차

Figure 1. ACD family Pedigree (ACD 유전병 가계도) .....	17
Figure 2. Vector 디자인 .....	27
Figure 3. ACD 환자의 임상적 소견 .....	31
Figure 4. ACD 환자의 Skin Biopsy(피부생검) 현미경 사진 .....	32
Figure 5. Candidate gene filtering 과정 .....	38
Figure 6. 돌연변이 생성 결과 전기영동 사진 .....	44
Figure 7. WT/Mut 플라스미드 염기서열 분석 비교 .....	45
Figure 8. <i>GPNMB</i> 발현 실험 .....	46

## 표 목차

Table 1. Sanger sequencing에 사용한 후보유전자들의 PCR primer sequence .....	24
Table 2. ACD 후보유전자들의 염기서열변이 목록 .....	35

# I. 서론

Genome이란 용어는, 유전자(gene)와 세포핵 속의 염색체(chromosome)의 합성어로, 염색체상에 존재하는 유전자들을 총칭한다. 인간의 경우 23쌍의 염색체 안에 대부분의 유전정보를 담게 되는데, 남자의 경우 상염색체 22쌍과 XY염색체, 여자의 경우 상염색체 22쌍과 XX염색체를 지니게 된다. 한 개체 내에 있는 모든 세포는 동일한 수의 염색체와 유전정보를 지니기 때문에, 하나의 세포만 분석하여도 전체 게놈정보를 파악할 수 있다. DNA상에 존재하는 유전정보는 A(아데닌), C(시토신), G(구아닌), T(티민) 네 종류의 염기들이 3개씩 조합된 것으로써, 유전암호가 특정 아미노산을 지정하고 이 아미노산들이 수십, 수백 개가 연결됨으로써 단백질이 형성된다. DNA를 모두 풀어놓으면 길이가 1.8m정도이며, 30억개 이상의 염기서열을 지닌다. 하지만 총 염기서열 중, 단백질 생산에 관여하는 '유전정보'를 가지는 염기쌍은 2%에 불과하다. 인간 생체 내에 기능을 수행하는 유전자는 약 2만개이며, 인간게놈의 염기서열은 거의 비슷하여 인종이 달라도 겨우 0.1%의 차이를 지니게 된다. 남녀간의 차이도 Y염색체 때문이지만 2%에 머무른다.

누군가에게 특정 병이 발병하는 데에는 유전적인 요인과 환경적인 요인이 작용한다. 그 중, 특정 유전자가 부모로부터 유전되어 발병하는 질환을 '유전병'이라고 한다. 유전병은 주로 DNA 염기서열 중 일부에서 돌연변이가 발생하고 그 결과 단백질 기능의 이상을 초래하거나 유전자의 발현양상의 변화로 인해 생기는 질환들이다. 염색체이상에 의한 질환(예: 다운증후군)은 특정유전자를 부모에게 물려받은 것이 아닌, 생식세포가 분열될 때의 염색체 non-disjunction이 원인이기 때문에 엄밀히 말하면 '유전병'의 범위에서 제외되기도 한다. 유전병은 유전자의 우열관계로 발병여부가 결정 되기도 한다. 우성 유전병일 경우, 우성 유전자를 1개만 가지고 있더라도 발병이 되는 반면에, 열성 유전병일 경우에는 열성 유전자를 2개 모두 가지고 있어야만 발병된다. 성

염색체 유전인지, 상염색체 유전인지에 따라서도 발병 경향이 달라지는데, 상염색체 유전일 경우 성별에 관계없이 비슷한 비율로 발병 가능성이 있다. 반면에 성염색체 유전일 경우 Y염색체 유전이라면 남자에게만 발병될 것이며, X염색체 유전이라면 X유전자를 1개만 가지는 남자에게 더욱 빈번하게 발병한다. 하지만, 돌연변이 유전자를 지니고 있음에도 해당 질병이 발병되지 않은 경우도 많다. 이를 침투도 (penetrance)라고 말하는데, 특정 유전자를 갖는 개체들 중 실제로 표현형을 지니는 개체의 빈도를 백분율로 나타낸 것이다. 다른 유전적인 배경이나 환경적인 요인으로 발병여부가 달라지며, 특정 유전자를 지닐 때 항상 발병되는 경우를 penetrance = 1 이라고 한다. 암, 당뇨, 고혈압 등의 호발성 질환들에 비해 희귀유전병의 경우 상대적으로 침투도가 높은 특징을 보인다. 그리고 유전병 발병원인이 항상 단 하나의 유전자 이상, 혹은 점돌연변이인 것은 아니다. 여러 가지 유전자 조합의 결과물로 혹은 여러 유전자 결과물들의 조합에 이상이 생겼을 수도 있고, 간접적으로 특정 pathway에서 중간 물질에 이상이 생겨 결과적으로 질병이 생길 수도 있다.

동계교배(Inbred marriage)가 이루어질 경우 혈연관계가 가까워 비슷한 유전자를 지니기 때문에 열성유전자들이 만날 확률이 높아지고, 그로 인한 질병이 발생할 확률 또한 높아진다. 우리나라를 비롯한 대부분의 나라가 가까운 혈연관계 내의 결혼을 금지하고 있지만, 파키스탄의 경우 문화적으로 동계결혼이 보편화 되어있고, 이 때문에 여러 열성 유전병들이 발병된다. 그리고 과거에는 특정 혈족, 순수혈통 가문을 유지하기 위해 계급 내, 가문 내에서만 결혼이 이루어지기도 했었다. 순수혈통을 지키기 위한 일들이 사실은, 질병을 일으키는 유전자들의 모임이 되면서 가계도내에서 열성유전병을 일으키게 된다. 대표적인 예가 유럽의 전역을 통치했던 ‘합스부르크 가문’인데, 5대에 걸쳐 동계결혼이 이루어졌으며, 대대로 물려진 심한 주걱턱에 ‘합스부르크의 턱(Habsburg’s chin/jaw)’ 라는 이름까지 지어졌다. 마지막 왕인 카를로스 2세는 신체적 불구뿐 아니라 정신지체도 있었으며, 결정적으로 불임이었기 때문에

더 이상 대를 이을 수 없었고, 합스부르크 혈통 보존을 위한 노력은 이렇게 막을 내리게 된다.

유전병들에도 정신질환, 안과질환, 암, 기형 등 많은 경우가 있는데, 본 연구에서는 그 중에서도 피부유전질환을 대상으로 설정하였다. 피부질환은 대부분 생명과 직결되지는 않지만, 질병의 종류에 따라 단순히 외관상으로 보여지는 피부이상부터, 가려움 혹은 예민성으로 인하여 생활이 불가능할 정도의 피부질환까지 다양하다. 대표적인 피부유전병으로는, 우선 백색증이라 흔히 불리는 알비노증(albinism)이 있다. 멜라닌 합성 결핍에 의한 피부, 털, 눈 모두가 백색으로 나타나게 되는 선천성 유전질환이다. 두 번째로는, 우리나라에서도 흔한 백반증(vitiligo)인데, 멜라닌 색소의 결핍 혹은 파괴로 인해 흰색반점 또는 패치형태의 흰색 얼룩이 생기는 질환이다. 원인에 대한 가설로는 ‘자가면역질환설’이 유력한데, 면역세포가 정상 멜라닌색소세포를 공격하기 때문이라고 추측된다. 미국의 유명 가수인 마이클 잭슨 또한 백반증 환자였다. 세 번째는 색소성 건피증(Xeroderma Pigmentosum)이다. 이 질환은 손상 받은 DNA를 재생하는 효소인 DNA Endonuclease의 부재로 나타나는데, 햇빛을 받아 DNA가 망가졌을 때 고칠 수 있는 효소가 없게 된다. 그렇기 때문에 햇빛에 매우 민감하고, 피부이상뿐 아니라 피부암까지도 발전할 수 있으며, 눈과 신경계까지도 이상이 생길 수 있다. 네 번째 피부유전질환의 예는 포르피린증이다. 헤모글로빈 합성에 필요한 전구물질인 포르피린이 8단계에 거쳐 적혈구로 생성되는데, 이 과정이 제대로 이루어지지 않아 포르피린이 그대로 신경계통이나 간, 피부에 과도하게 침착되는 질환이다. 햇빛에 과민반응 하게 되며, 피부에 물집이 생기고 색소침착으로 눈이 따갑게 느껴진다. 상대적으로 입술과 잇몸이 눈에 띄게 우묵하게 들어가 송곳니가 튀어나온듯한 외모, 뱀파이어상으로 보여지기도 한다.

본 연구에서는 피부유전질환 중에서도 ACD (Amyloidosis Cutis Dyschromica)의 원인유전자를 규명하는 것을 연구목표로 설정하였다. ACD는

아밀로이드의 침착으로 인해 피부에 과색소와 저색소증이 같이 일어나면서 피부 전반에 걸쳐 반점이 생기는 질환이다. 열성유전질환 중 하나로, 세계에서 48명만이 이 질환을 지닌 것으로 보고된 매우 희귀한 질병이다. 열성 유전병이기 때문에, 가계도 내의 자녀세대 중 환자들은 질병을 일으키는 돌연변이 열성 유전형질을 2개 가질 것으로 예상된다. 그리고 부모의 경우, 정상 형질의 우성 유전형질과 돌연변이 열성 유전형질을 각각 1개씩 가질 것으로 예상된다. 그리고 자녀세대 중 이 질환이 없는 정상인은 정상 유전형질 2개 혹은 정상 유전형질과 돌연변이 유전형질을 각각 1개씩 가질 것으로 예상되며, 이 가설을 기반으로 원인유전자를 규명해 나가고자 하였다.

특정 질병에 대한 후보유전자를 규명하는 일은 단순히 유전자를 찾아내기 위한 일은 아니다. 모든 연구의 궁극적인 목표는 사람의 생명을 유지시키고 더 나은 삶을 살아가기 위해 도움을 주기 위함이다. 원인유전자를 규명했을 때에, 첫 번째로는 예방의 가능성이 있다. 현재도 산전진단을 위해 염색체 이상검사가 빈번히 이루어지고 있으며, 특정 질병에 대해 자손에게 유전이 되었는지, 혹은 본인이 그 질병에 대한 발병 위험성을 가지고 있는지에 대한 여부를 유전자 검사로 쉽게 알아낼 수 있다. 두 번째로는, 치료에 목표를 두는 연구이다. 원인 유전자 규명을 통해 질병의 발병 기작을 유추해낼 수 있고, 후속 연구를 통해 치료방안을 찾아낼 수 있을 것이다. 혹은 직접적인 유전자 조작을 통해 유전자 치료도 가능해질 수 있다.

## II. 이론적 배경

### 1. Amyloidosis Cutis Dyschromica

약칭으로 ACD라고도 부르며, 각질형성세포(keratinocyte)에서 비롯된 아밀로이드(amyloid)가 피부에 침착되는 병이고, Primary cutaneous amyloidosis (PCA)의 일종이다. PCA는 어떠한 다른 몸의 이상 없이 아밀로이드라는 전분류가 피부에 침착 되는 질환이다(Madarasingha *et al.*, 2010). Oncostatin M receptor과 연관이 있으며 반상유전분증(Macular amyloidosis), 유전분증태선 (Lichen amyloidosis), 결절아밀로이드증 (Nodular amyloidosis) 세 가지로 나뉘어진다(Bourke *et al.*, 1992; Schreml *et al.*, 2010). 현재 한국어로 규명된 명칭은 없다.

Amyloidosis란 ‘아밀로이드증’으로 희귀 난치성 질환이며 단백질의 형성 과정에서 형태에 이상이 생겨 여러 장기와 조직에 섬유질이 형성되는 질환이다. ACD 환자들의 대부분은, 유전적 요인 없이 환경적 요인으로만 개별적으로 발병하는 경우보다는, 가족 내에서 유전적인 요인으로 인해 질병이 발병하고 대부분의 환자들은 동아시아인들인 것으로 보고되었다. 가족 내에서 발병하는 경우 평균 발병나이는 6세였고, 개별 발병 경우에는 평균 발병나이는 23세이다(Mahon *et al.*, 2015). 피부 이외에 골수종, 마크로글로불린혈증, 만성염증, 탈모, 결절 등을 일으키는 전신성 아밀로이드시스(systemic amyloidosis)는 수반되지 않고, 가려움증만 동반되는데 이 또한 환자 중 16%만이 가진다(Mahon *et al.*, 2015). 아시트레틴 (Acitretin; 레티노이드 종류 중 하나; 레티노이드는 비타민 A와 화학적으로 연관된 물질군으로 레티노산, 레티놀, 레티알데히드를 모두 포함)을 처방한 10명의 환자 중 7명이 상태가 호전되었다고 보고된 바 있다(Ozcan *et al.*, 2005).

ACD는 황반(macular)이나 그물 모양(reticulate)의 과색소침착(hyperpigmentation)과 대칭적인 방울(guttate) 모양의 저색소침착(hypopigmentation)이 함께 나타난다(Mahon *et al.*, 2015). 임상적인 관찰로는 선천성 이상각화증(Dyskeratosis Congenita)이나 색소성건피증(Xeroderma pigmentosum)과 매우 유사해 보이지만, 피부생검(Skin biopsy)을 대상으로 Congo Red 염색 후 현미경으로 관찰할 때, 유두진피(papillary dermis)에서 아밀로이드의 침착이 발견된다는 점이 다른 질병들과 확연히 구분되는 특징이다(Fernandez-Flores, 2012; Ho *et al.*, 2009; Ogino and Tanaka, 1977). 주로 몸통부분부터 시작해 점점 사지로 번져나가는 추세이다.

Morishima그룹이 수행한 연구를 통해 1970년대에, 일반적인 PCA와는 차별되는 독특한 임상양상이 관찰되었으며 이를 통해 ACD라는 병이 처음으로 알려지게 되었는데, 그 특징들은 세 가지가 있다. (i) 색소변이가 있는 점 혹은 그물모양의 반점들이 관찰된다는 점, (ii) 가려움증이 약하거나 없음, (iii) 조직학적인 측면에서의 관찰에서, 표피에서 가까운 유두진피에 아밀로이드의 침착이 관찰된다는 점, (iv) 주로 사춘기 즈음에 발병한다는 점이다(Morishima, 1970). 처음 ACD를 규명하게 된 1970년 1월부터 2014년 9월까지 학계에 보고된 ACD 환자들은 총 48명으로, 전세계적으로 매우 희귀한 질병이다. ACD는 다른 수반되는 증상이 없고, 천천히 점차적으로 더 많은 반점들이 생기기 때문에 환자들이 질병임을 인지하고 진단을 받기까지 매우 오랜 시간, 보통 10여년이 지난 후 진단을 받게 된다(Mahon *et al.*, 2015).

## 2. 멘델의 열성유전 모델

사람을 비롯한 대부분의 생명체들은 2배체로써 같은 위치에서 두 가지 대립형질을 가지고 있는 경우에 표현형으로 나타나는 형질이 ‘우성’, 나타나지 않는 형질을 ‘열성’ 이라고 부른다. 한 사람이 가지고 있는 염색체 쌍에 두 가지 형질이 모두 열성일 경우에만 열성 표현형이 나타나게 된다. 유전자가 상염색체(1~22번 염색체)에 있는지 성염색체(X,Y염색체)에 있는지에 따라, 관련 질병들도 i) 상염색체 우성 유전병, ii) 상염색체 열성 유전병, iii) 성염색체 우성 유전병, iv) 성염색체 열성 유전병으로 나눌 수 있다. 사실 특정 표현형을 가지기 위해서는 단 하나의 유전자 변이 뿐 아니라, 환경, 혹은 여러 유전자의 상호작용, 우성/열성 표현형이 아닌 중간 표현형 등등 다양한 형태로 존재할 수 있지만, 1개의 유전자에서 비롯된 우성/열성 형질의 만남으로 2가지의 표현형이 나타나는 경우를 ‘Mendelian inheritance’ 라고 한다. 같은 형질을 2가지 지닌 경우를 ‘homozygous(동종접합)’ 이라고 하고, 다른 형질 2가지를 모두 가지고 있는 경우를 ‘heterozygous(이종접합)’ 이라고 한다.

회귀유전병 가계도상에서 부모와 자녀세대의 질병 발현 양상을 보고, 이 질병이 어떤 유전패턴을 따르는지 예측할 수 있다. 상염색체 우성 유전패턴인 경우, 성에 상관 없이 세대별로 모두 한 명 이상 환자(affected individual)이 있다. 반면에 상염색체 열성 유전패턴인 경우, 우선 성별에 연관성은 보이지 않으며, 정상인 부모에게서 돌연변이 표현형을 가지는 자녀가 태어나는 양상을 보인다. 이 경우, 정상인 부모는 둘 모두 정상 형질 1개와 돌연변이 형질 1개를 가지고 있는 ‘보인자(carrier)’ 라고 생각할 수 있다. 물론 부모에게서 없던 표현형이 자녀세대에서 나타나는 경우만을 열성유전패턴이라고 할 수는 없다. 예를 들면, 다지증(polydactyly)은 우성형질인데, 보통사람들에게는 많이 나타나지 않는다. 이 경우, 보통사람들이 5개의 손가락을 만들어내는 열성유전자를

보편적으로 가지고 있음을 알 수 있다. 어떤 형질을 정상으로, 돌연변이로 보느냐에 따라 유전패턴을 규정지을 수 있다.

하나의 돌연변이 유전자의 존재여부로 같은 표현형이 나오리라는 법은 없다. 어떠한 표현형으로 나타나기까지 영향을 줄 수 있는 요인들이 많다. i) 환경 (environment), ii) 침투도 (penetrance; 어떠한 유전자를 갖는 개체의 집합 중 표현형으로 나타난 개체들의 빈도를 백분율로 나타내는 것), iii) 표현도 (expressivity, 개체 별 표현형의 심한 정도), iv) 상위성 (Epistasis; 어떠한 유전자가 다른 유전자의 표현형에 영향을 주는 현상), v) 다면발현 (pleiotropy; 한 유전자가 두 가지 이상의 형질 발현에 관여하거나 영향을 주는 현상), vi) 이질성 (Heterogeneity; 다른 유전자가 동일한 표현형을 일으키는 현상, 한 가지 질병의 원인유전자가 여러 개인 경우) 등이 있다.

### 3. 유전학연구 측면에서 동계결혼의 이점

혈육관계가 가까울수록 비슷한 유전자를 가질 확률은 높다. 일반적인 경우 세대를 거치게 되면서, 사회적 선택과 우/열성 표현형의 선택 등등의 이유로 '정상'의 유전자들이 전해지지만, 누구든 질병을 일으킬 수 있는 열성유전자들을 지니고는 있다. 하지만 열성유전자보다는 질병을 일으키지 않는 우성유전자를 지닐 확률이 훨씬 크기 때문에 열성유전자를 전해주더라도 표현형은 우성형질로 나타날 확률이 크다. 그런데, 가까운 혈육관계에 결혼을 하게 되는 경우는 다르다. 보인자인 양쪽 부모로부터 열성유전자를 각각 하나씩 유전 받은 자손이 태어날 확률이 25%이므로 열성유전병이 발병될 확률이 커진다. 이 때문에 우리나라를 포함한 여러 나라들이 동계간의 결혼을 법으로 무효화하고 있다. 우리나라의 경우, 과거에는 '동성동본 내 혼인금지'였지만 현재는 '8촌 이

내의 혈족 사이에서의 혼인을 무효화(민법 제809조제1항)' 를 법으로 지정하고 있다.

하지만 파키스탄의 경우에는 다르다. 그들에게는 3촌 이내의 결혼이 금지되어 있지만, 4촌 혹은 6촌 등등 친족간의 결혼은 오히려 보편화 되어있다. 사우디아라비아에서는 80%, 파키스탄의 경우 70%가 친족간의 결혼을 할 정도로 매우 흔하고, 사실상 사촌간의 결혼을 더 선호하는 편이다. 이런 풍토는 그들의 뿌리깊은 문화에서 그 이유를 찾아볼 수 있다. 빈부격차로 인한 계급을 유지시키려는 목적도 있었고, 부족주의 시대에 타부족을 불신했던 관습이 아직까지 남아있어 동계결혼이 만연하다. 그래서 현재 유전적인 질병을 가지는 사람들이 매우 많고, 동계결혼으로 발생하는 희귀 유전병 가계도를 이용하여 연구를 진행하였다.

#### 4. 돌연변이의 종류

염기서열 변화로 실질적으로 아미노산에 변화가 생기는 것을 Nonsynonymous Mutation 라고 한다. 반대로, Synonymous Mutation은, 염기서열에 변화가 있기는 했지만 코딩되는 단백질은 같을 경우, '동의코돈간의 변이' 라고 하며 '잠재성 돌연변이(silent mutation)' 라고도 한다.

Nonsynonymous Mutation의 예로써는, i) Nonsense mutation : 정지 돌연변이, 염기서열에 변화로 정지코돈(UAG, UAA, UGA)가 생기는 경우, 전체 단백질이 아닌 중간에 끊긴 단백질을 형성하게 된다. ii) Missense mutation : 과오 돌연변이, 염기치환으로 아미노산에 대응하는 코돈이 바뀌고, 그 결과 활성이 변하거나 소실된 단백질이 합성된다. iii) Frameshift mutation : 틀이동 돌연변이, DNA 상에서 3의 배수가 아닌 소수 개의 염기가 삽입 혹은 결실된 경

우, 3배수로 읽혀지는 코돈의 순서가 밀리게 되면서 돌연변이가 생긴 위치 이하의 아미노산들에 모두 변화가 생기게 된다.

실제 아미노산에 이상이 생겼기 때문에 비정상적인 결과로 유전병이 발생하게 된다. 본 연구 주제인 ACD 또한 아미노산에 생긴 이상이 원인이라 가정하고, Nonsynonymous Mutation에 한하여 스크리닝하도록 한다.

## 5. 생물정보학을 이용한 염기서열 분석

### (1) 생물정보학(Bioinformatics)

Bioinformatics는 생물정보학 이라고 하며, 생명의 본질인 DNA 에 담겨있는 유전정보를 효율적으로 분석하기 위하여 전통적인 유전학과 컴퓨터 공학분야의 융합을 통해 생성된 학문이다. 하지만 단지 DNA에 국한되지 않고, 여러 가지 생명현상에 대한 의문점을 해결하는데 있어서 컴퓨터와 디지털 시스템을 이용하는 모든 분야를 일컫는다. 최근 유전정보들을 데이터베이스화 할 수 있게 되면서 매우 중요하면서도 각광받는 분야가 되었다. 방대한 규모의 유전체(genome)정보를 컴퓨터와 소프트웨어를 이용해서 염기서열 데이터를 정리하고 분석한다. 인체의 유전자 정보가 저장되어 있는 바이오 칩을 이용할 경우 개인의 유전자 변이나 이상여부를 쉽게 알 수 있어, 질병의 조기진단이 가능해진다. 인간의 혈액세포로부터 추출된 DNA 샘플을 sequencing을 통하여 문자화된 서열정보로 얻어내는 과정도 모두 생물정보학을 이용하는 과정이다.

생물정보학에서 주로 다루는 대상은 유전체(Genome)과 유전자(gene), 단백질(Protein)이다. 사실, 어떠한 생명체를 유전자 수준의 연구에서 모든 것을 이해한다는 것은 불가능하다. 그렇기 때문에, 각 요소간의 상호작용이 중요하다.

그렇기 때문에, 하나의 생명체를 하나의 네트워크로 인식하고, 어떤 관계를 맺고 있는지 연구하는 오믹스(체학;Omics) 수준에서의 접근방법이 매우 중요하다. 이 범주에 포함되는 분야에는 Genomics (유전체학), Epigenomics (후성유전체학), Structural Genomics (구조유전체학; 유전체 정보를 통해 발견되는 단백질의 3차구조와 기능연구), Glycomics (글리코믹스; 당 사슬(glycom) 연구), Transcriptomics (전사체학), Proteomics (단백질체학), Metabolomics (메타볼로믹스, 대사체학; 세포과정(Cellular process) 중 발생하는 대사산물에 대한 종합연구) 등이 있다. 이 모든 분야를 잘 아우르면서 서로 정보교환이 원활해지기 위해서는 생물정보학의 데이터베이스 형성이 매우 중요하다. 각각의 연구분야에서 나오는 결과물인 막대한 양의 자료와 정보들을 온라인상의 데이터베이스에 추가되면서 더욱 다양한 정보들이 축적된다.

생물정보학을 이용하여 유용하게 쓸 수 있는 정보 분석방법 중 한가지가 'BLAST' 이다. BLAST는 'Basic Local Alignment Sear Tool' 의 약자로, 아미노산배열 혹은 DNA 염기서열을 비교할 수 있는 알고리즘이다. 사람을 비롯한 여러 종들의 게놈정보가 데이터베이스에 등록이 되어있다. 사용자가 직접 Sequencing과 alignment를 거쳐 생성된 염기서열이 어떤 종의, 어느 염색체의 어떤 부분에 해당하는 것인지 문자화된 서열정보로 비교가 가능하다. BLAST 보다 감도는 떨어지지만 더 빠른 속도로 비교 가능한 BLAT(Basic Like Alignment Tool)도 있다. 그리고 단백질의 측면상(profile)을 이용하여 관련단백질을 검출할 수 있는 검색도구로서 PSI-BLAST(Position-specific Iterative BLAST) 도구도 있다.

## (2) 차세대 염기서열 분석 (Next-Generation Sequencing)

차세대 염기서열 분석(Next Generation Sequencing)은 다양한 DNA 염기서열분석(DNA Sequencing) 기술들 중 하나로 DNA 또는 RNA 염기서열을 고속, 대용량으로 해독할 수 있는 방법이기에 ‘Massive parallel sequencing’, ‘High-throughput sequencing’ 이라고도 부른다. 13년이 걸려 2003년 인간의 게놈프로젝트를 완성한 것에 비해, 현재는 4주면 한 사람의 모든 유전정보를 얻을 수 있다. 시간이나 비용 및 응용성에서 전통적인 Sanger sequencing, PCR, Microarray 기술을 대체할 수 있을 정도로 그 활용도가 매우 높다.

NGS는 한 작업을 동시에 수행하는 병렬 컴퓨팅과 유사하게, 유전체를 무수히 많은 조각으로 나눈 뒤 각각의 염기서열을 조합하여 유전체를 해독하는 분석 방법이다(Meder *et al.*, 2011). Illumina 회사 장비를 이용한 NGS 실험 과정은 다음과 같다.

- ① 샘플 준비(Sample Preparation): 혈액, 타액, 구강 상피세포 등에서 추출된 DNA 혹은 RNA
- ② Library 제작: 염기서열 분석이 가능하도록 초음파를 이용하여 무작위로 DNA 절단, 절편의 끝을 변형시켜 Adapter(인식이 용이하게 하는 인식표)를 부착(ligation)한다
- ③ Library 정량: DNA polymerase를 이용하여 Library를 증폭시켜 Flow cell 내부에 Cluster(DNA 절편이 올리고와 인접한 올리고에 브릿지를 형성해 이중 가닥을 만들어내는 과정을 수없이 반복하여 cluster, local clonal DNA colonies를 만듦)를 형성한다
- ④ Sequencing: Hiseq이라는 해독기를 이용, 해독을 위한 primer가 adapter sequence에 부착, 형광 표지 되어있는 A,T,G,C 염기가 sequence에 맞추어 하나씩 부착되며 촬영한다.
- ⑤ 데이터 분석(Data analysis): 분석 소프트웨어를 이용, 대부분 pair-end 를

이용하기 때문에 양방향으로 시퀀싱 수행. 미리 알려진 reference sequence와 비교하여 해독된 염기서열을 정렬(Alignment), 대조군과 비교하여 변이(variation)의 위치와 유형을 발견한다.

DNA 염기서열 정보로부터 얻을 수 있는 가장 핵심적인 부분은 집단 안에서의 개인별 차이를 보이는 변이이며 i) SNP (Single Nucleotide Polymorphism), ii) MNP (Multi Nucleotide Polymorphism), iii) InDel (Insertion & Deletion), iv) Structural Variation (Inversion, Translocation), v) CNV (Copy Number Variation) 등이 있다.

전체 genome sequencing을 하는 것에 비해, 특정 부분의 유전자 영역에만 집중하여 시퀀싱할 경우 비용을 감소시킬 수 있는데, Targeted genome sequencing 이라고 한다.

NGS는 더 빠르게, 더 저렴하게 진행될 수 있도록 현재에도 기술 개발 중에 있다. NGS라는 기술로 DNA를 신속하고 정확하게 해독해냄으로써 인간 생명 현상을 밝혀, 예방의료 및 맞춤의료분야를 통해 인류의 보건복지 향상에 크게 기여할 것으로 예상된다.

### (3) 엑솜 염기서열 분석법 (Whole-exome sequencing)

엑솜 시퀀싱(Whole Exome Sequencing)이란 intron, promotor, UTR 영역을 제외하고 단백질을 암호화 하고 있는 엑손(protein-coding region)들만 염기서열 분석하는 방법이다. 인간의 게놈상에는 약 180,000개의 엑손(exon)이 존재하며, 이는 사람의 총 염기서열의 1%에 해당한다.

이런 방식은 Whole-genome sequencing까지 하지 않고도 저비용으로, 멘델의 유전법칙을 따르는 유전병과 일반 질병들의 유전적 원인을 규명할 수 있다는 큰 장점이 있다. 그 중에서도 '멘델 법칙을 따르는 희귀 유전병'의 유

전적 원인 분석에 가장 적합하다. 그 이유는 첫 번째로, SNP array를 할 경우 대다수가 공유하고 있는 유전적 변이만을 알 수 있는 데 반해, 엑솜 시퀀싱은 희귀 유전병을 가지는 극소수의 모든 유전정보를 알 수 있기 때문이다. 그리고 두 번째로는, 대부분의 질병을 일으킬만한 심각한 변이들은 대부분 단백질을 암호화하고 있는 엑손 부위에 변이가 일어나는 경우이기 때문이다.

엑솜시퀀싱의 진행방식은 다음과 같다. 우선 전체 게놈 중 엑손 부위만 추려내고 그 후로는 NGS 진행방식처럼 flow cell 내에 Cluster를 만들고 시퀀싱이 진행된다. 엑손 부위만 걸러내는 방식들, Target-enrichment strategies로는 4가지를 예로 들 수 있다. 많은 target-enrichment strategy들이 2005년에 본래 명명된 'Direct genomic selection(DGS)' method로부터 유래되어 발전해왔다.(Bashiardes *et al.*, 2005)

a) PCR : Polymerase chain reaction. 특정 유전자 부위 증폭을 위해 일반적으로 쓰는 실험 방법 중 하나인데, 엑솜시퀀싱에서도 유용하게 이용될 수 있다. 특히 각각의 엑손 부위를 증폭시킬 수 있는 primer를 동시에 여러 개를 넣음으로써, 동시다발적으로 각 염색체 내의 엑손에 해당하는 부위들을 증폭해낼 수 있다.

b) Molecular inversion probes (MIP): 특정 영역의 양 끝에 붙는 단일 가닥 DNA 올리고뉴클레오티드(single stranded DNA oligonucleotide)를 probe로 이용한다. 특정 부위의 양쪽에 붙어 원모양을 이루는데(target circularization), 이 때 원 모양을 이루지 못하고 linear로 남아있는 probe들은 exonuclease에 의해 제거된다.(Mertes *et al.*, 2011; Turner *et al.*, 2009)

c) Hybrid Capture: 특정 염기서열을 가지고 있는 Microarray를 이용한다. Genomic DNA는 double-stranded fragments로 되어 adaptor가 부착되고, microarray위의 oligo들과 hybridization이 일어나, 선택적으로 PCR이 진행된다.(Turner *et al.*, 2009)

d) In solution capture: 합성된 target-specific probe들을 풀어놓고, 용액 내에서 조각난 DNA와 probe들이 hybridization된다. 선택적 특이성을 지닌 probe들이 beads와 결합하게 되고, probe와 더불어 결합된 DNA가 pull down 된다.

## 6. 희귀유전질환

‘희귀병’에 대한 정의는 사실 명확하게 정의되어 있지 않다. 각국의 보건 정책에 따라 모두 다르게 정의되지만 보통, 전체 인구의 0.04~0.06% 이하에 계서 나타나는 병을 의미한다. 하지만 확률을 만족하더라도 치명적이지 않거나, 만성적이지 않고 치료가 가능하다면 희귀병으로 정의하지는 않는다. 사실 어떤 집단에서냐에 따라 그 빈도수가 다양하게 나타나기 때문에, 어떤 지역에서는 희귀병이라 할지라도 다른 지역에서는 그보다 보편적인 질병일 수도 있다.

현재 희귀병으로 분류되고 있는 병들은 5,000에서 7,000가지에 달한다. 이 중, 약 400가지만 치료방법이 알려져 있다. 희귀, 난치성질환의 유병인구는 매우 적기 때문에 원활한 연구가 이루어지기 어렵다. 때문에 발병 원인이 정확히 규명되지 않아 치료방법과 대체의약품이 개발되기 쉽지 않다. 그렇기 때문에 희귀병의 치료방안도 없고, 연구를 위한 투자도 잘 이루어지지 않는 ‘orphan disease’라고 부르기도 한다.

### III. 연구 방법

#### 1. 연구 대상

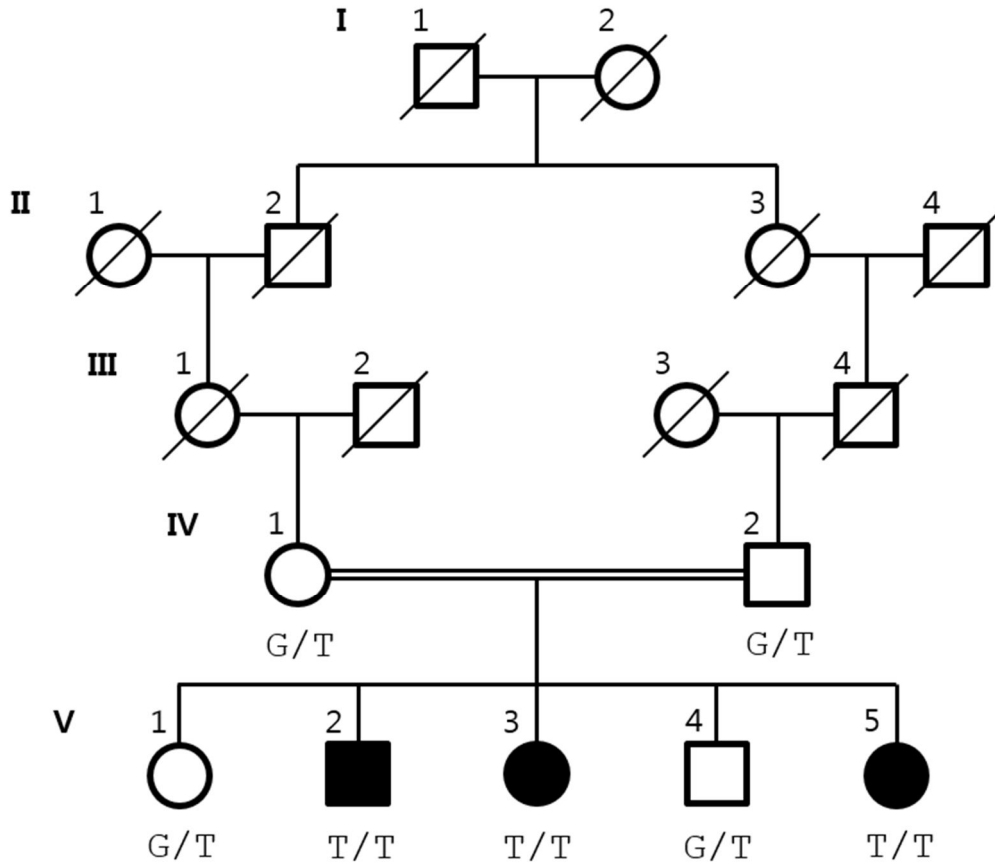
본 연구에서는 희귀유전병 가계도를 확보하기 위하여 파키스탄의 카이버의 과대학교 (Khyber Medical University)내 Musharraf 교수 연구팀과 국제공동연구의 일환으로 진행하였다. 파키스탄의 동계결혼 및 다자녀 문화는 가계도에 기반한 유전병 연구에 큰 장점으로 작용한다.

가계도 확보 노력결과, ACD라는 희귀한 피부유전병을 지니고 있는 한 가족이 발견되었다. ACD는 상염색체 열성유전병이기 때문에 부모는 정상적인 피부를 가지는 정상인임에도, 자녀들 5명 중 3명이 피부색소이상을 지니는 피부병 환자이다(Figure 1).

환자들의 양쪽 부모는 증조부모가 같은 6촌지간이었다. 부모와 5명의 자녀, 총 7명이 연구에 응해주었고, 7명 모두의 혈액샘플을 채취하였다. 부모세대(IV)의 윗 세대들은 모두 사망했기에 그 이상의 혈액샘플은 얻을 수 없었지만, 그 누구도 ACD와 관련된 표현형을 지니는 사람은 없었다.

자녀들은 총 5명으로 2남 3녀였다. 그 중, 딸 1명과 아들 1명은 정상인이고, 딸 2명과 아들 1명은 ACD 환자였다.

Figure 1. ACD family Pedigree (ACD 유전병 가계도)



IV세대에서 1번, 2번 구성원간의 결혼은 동계결혼이므로 이중선으로 표시하였다. *GPNMB* 유전자 염기서열에서 c.1124G>T변이는 p.Gly375Val를 초래한다. 가계도상에서 환자들은 모두 *GPNMB* c.1124G>T 돌연변이를 2개 보유하고 있다.

## 2. 가설

Amyloidosis Cutis Dyschromica는 이미 상염색체 열성유전병으로 알려져 있다. 하지만 이 질병과 관련된 원인유전자는 아직 보고된 바 없다.

열성유전병이기 때문에, 부모세대는 우성인 정상 유전자를 1개씩, 열성인 원인 돌연변이 유전자를 1개씩 가지고 있는 보인자(Carrier)(유전자형: WT/Mut)일 것이다. 자녀세대에서 ACD 환자의 경우, 열성인 원인 돌연변이 유전자를 2개 모두 가지고 있을 것(유전자형: Mut/Mut)으로 가정한다. 자녀세대 중 정상인인 3명의 유전자형은 확정 지을 수는 없다. 하지만 적어도 1개 이상의 우성 정상 유전자를 가지고 있을 것이다. 유전자형은 WT/WT 혹은 WT/Mut 일 것으로 추측된다.

환경적 요인이나 여러 유전자의 복합적인 상호작용의 결과로 나오는 Complex disease가 아닌, 단 하나의 돌연변이 유전자로 인한 질병일 것이라고 가정한다. 즉, 멘델의 유전법칙을 따르는 전형적인 Mendelian disorder 일 것으로 가정한다.

상기 기준들을 만족시키는 염기서열변이들을 후보 돌연변이로 간주하였다.

## 3. 연구방법

### (1) DNA 정제 및 엑솜 캡처링

G-DEX™ IIb Genomic DNA Extraction Kit [for Blood] 를 이용하였다. 0.1~20 ml 정도의 혈액샘플로부터 약 4.6~813  $\mu$ g의 gDNA를 추출할 수 있

다. 혈액 내 백혈구 안에 있는 DNA를 추출해내며, 혈액 5 ml당 백혈구(WBC; white blood cell)가 약 4천만개 정도 있다.

진행 과정은 「용해(Lysis) → 단백질 제거(protein removal) → DNA 침전(precipitation) → DNA 수화(hydration)」 순으로 이루어진다.

파키스탄으로부터 가계도 구성원들의 혈액, 3~10 ml을 EDTA Vacutainer 튜브에 담긴 상태로 전달 받은 후 G-DEX™ kit을 이용하여 DNA를 추출해낸 후 -80 °C 혹은 -20 °C에 보관하여 사용하였다.

Exome Capturing에는 SureSelect V4 kit (Agilent Technologies, Santa Clara, CA, USA)를 사용하였다. Exome Sequencing에는 100-bp paired-end sequencing으로 Illumina HiSeq2000 platform (San Diego, CA, USA)을 사용하였다.

## (2) DNASTAR Lasergene 12 core suite program

NGS가 진행되어 나온 ACD family 내 개개인의 genome 정보가 담긴 fastq 파일을 받게 되면, DNASTAR사의 Lasergene 12 Core suite 프로그램을 이용하여 후보유전자들을 추려내게 된다.

여기서는 첫 번째로, fastq file 형태의 염기서열 정보를 alignment 시켜주는 assembly 과정이 필요하다. 여기서 쓰이는 프로그램이 DNA Seqman Ngen 12이다. Human reference genome sequence (Homo sapiens GRCh38\_dbSNP142)에 대조시켜 NGS data (fastq files, genome templates) 들을 assembly 시킨다. 주로 이용한 것은 reference sequence와 비교하며 alignment하는 'Templated assembly' 였지만, reference 없이 가지고 있는 파일들만 조합해서 만드는 *de novo* assembly도 가능하다. 1개 혹은 여러 개의 샘플을 동시에 assembly할 수 있다.

두 번째로는, Assembly가 완료되어 개개인의 염색체 별 genome정보가 일렬로 정돈된 정보를 이용하여, 직접적으로 filtering을 시작하는 작업이 필요하다. 이 때 사용하는 프로그램이 DNASTAR ArrayStar12이다. Large scale variant analysis에 이용된다. Assembly가 완료된 NGS data를 기반으로 가계도 내 유전패턴을 기반으로 조건을 줄 수 있다. 여러 가지 조건에 부합하는 돌연변이 유전자, SNP들만 추려진 결과를 얻을 수 있다.

세 번째로는, 여러 개의 sanger sequencing 정보를 한꺼번에 열어, 그래프를 동시에 비교하는 프로그램이다. SeqMan Pro라는 프로그램이 쓰이며, Sanger sequencing의 assembly 혹은 분석, 혹은 NGS 결과로 assembly 완료된 파일을 열어 alignment된 자료들을 문자로 정렬된 염기서열로 비교할 수 있다. Read depth를 직접 눈으로 확인할 수 있으며, 같이 assembly가 진행된 여러 샘플들을 한꺼번에 비교할 수 있다.

### (3) 유전정보 검색에 유용한 데이터베이스

앞서 기술한 filtering 과정을 거치게 되면, 후보유전자들의 염기서열변이 목록이 추려진다. 후보유전자들의 이름과 돌연변이가 존재하는 위치까지 모두 알 수 있다. 이렇게 나온 후보유전자들의 이름을 손쉽게 검색하여 지금껏 연구된 바들로 특정 유전자에 대한 정보들을 쉽게 얻을 수 있도록 준비되어있는 몇 가지 database 사이트가 있다. 전세계에서 연구되는 정보들을 업로드 하여, 누구든지 정보를 받아 연구에 참고할 수 있다.

첫 번째로는, NCBI([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)). National Center for Biotechnology Information (미국 국립생물공학정보센터)사이트이다. 주로 SNP정보를 검색할 때 이용한다. 생물학 및 의학 분야의 방대한 데이터 분석을 위한 DB를 구축하고 분석도구를 개발하는 것이 주요임무이다.

두 번째로는, PubMed([www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed))이다. 생명과학 및 의학 논문 인덱스의 데이터베이스이다. 유전자 이름, 특정 질병, 혹은 연구자가 검색하고 싶은 몇 가지 query들을 넣어 검색하면 관련된 논문들을 모두 찾아볼 수 있다.

세 번째로는, OMIM([www.omim.org/](http://www.omim.org/)) (Online Mendelian Inheritance in Man)이다. 사람의 유전자와 유전 질병에 대한 온라인 목록이라고 할 수 있다. 특정 유전자 이름을 검색하면, 유전자이름의 뜻, 지금껏 보고된 특성들, 관련된 유전질병, 보고된 돌연변이, 분자유전체학 측면에서의 특징들 등 여러 가지가 잘 정리되어 나열되어있다. 유전자에 대한 전반적인 이해를 위해 유용하다.

네 번째로는, UCSC genome browser(<http://genome.ucsc.edu/>)인데, University of California, Santa Cruz에서 작업이 이루어지는 사이트이다. 기준 염기서열 (reference sequence)의 정보를 모두 담고 있다. 염색체 위치로 검색, 유전자로 검색, 동형 단백질(isoform)별 염기서열까지도 검색이 가능하다. 종 별 보존 정도(conservation) 파악 가능하며, 기준 염기서열(gDNA)와 실제 암호화되는 cDNA(complementary DNA, 상보적 DNA)서열, 아미노산서열, mRNA서열까지도 파악 가능하다.

다섯 번째는, 1000 genome data base (<http://browser.1000genomes.org/>)이다. 사람의 유전적 돌연변이에 대한 더 자세한 자료들이 있다. v76 GRCh37 버전에 기반하며, 26개 인종, 2504명에 대한 게놈상의 염기서열 변이 정보를 가지고 있다. 1000 게놈프로젝트는, SNPs와 구조적 돌연변이 그리고 그들의 haplotype 정보 등등, 사람의 유전적 다양성을 파악하기 위해 범세계적으로 진행된 합작이다. 25가지 인종 2500명의 게놈정보를 NGS 기술을 이용하여 염기서열을 알아냈다. 주로 인종별 Allele Frequency를 기입하기 위해 이용한다.

#### (4) MAF (Minor allele frequency) 조사

인간의 게놈에는 약 천 만개의 염기서열 변이들이 존재한다. 이들 변이들에 대한 모든 대립유전자 빈도(allele frequency)가 보고되어있지는 않지만, SNP(Single Nucleotide Polymorphism)이 보고된 경우를 비롯해 많은 경우가, 인종별로 특정위치에 대립유전자를 어떤 빈도로 지니고 있는지 자료조사 되어 공개되어있다. 이 경우에 '1000 genome browser' 혹은 'ExAC(Exome Aggregation Consortium)' 홈페이지를 이용하여 검색해볼 수 있다. GPNMB (p.Gly375Val)를 제외한 나머지 7개 후보유전자(PHRF1 p.Arg1046Gln, RASSF7 p.Arg223Gln, KIF21A p.Glu1201Asp, SLC25A22 p.Leu150Val, PKP3 p.Thr459Met, NUB1 p.Rrg9Cys, SOAT2 p.Thr254Ile)는 돌연변이 위치에 이미 SNP이 보고되어있었고, 그 dbSNP ID로 검색하여 대립유전자빈도를 알아낼 수 있었다(Table 2). 유전자 내 염기서열 빈도는 CEU(Northern and Western European)과 CHB/S(Han Chinese in Beijing/ Southern Han Chinese)와 YRI (Yoruba in Ibadan, Nigeria), 그리고 PJJ(Punjabi in Lahore, Pakistan)까지 조사하였다.

연구대상 질환인 ACD의 경우 전세계에서 48명만 보고된 희귀 열성 유전병이다. 그렇기 때문에, 일반사람들은 거의 ACD를 일으키는 돌연변이 형질을 지니지 않을 것이다. 특정 위치에서 Minor allele frequency(MAF)가 0.05라고 나와있을 경우, 20명 중에 1명이 그 열성 형질을 가진다는 의미를 지닌다. 열성유전병일 경우 이 열성 형질을 2개 지녀야 그 질병을 지니게 된다. 그러므로 MAF가 0.05인 경우  $0.05 \times 0.05 = 0.0025$ , 즉 10,000명 중에 25명, 400명 중에 1명꼴로 해당 열성유전병을 가진다고 할 수 있다. 그렇기 때문에 MAF를 파악하는 것이 중요한데, MAF가 0.00에 가까울수록 좋지만 후보유전자로 추려낼 때는 조금 범위를 넓혀 0.10 이하인 것으로 추린다.

## (5) 파키스탄 인종 내 유전자형 분석

본 연구의 대상 질병은 ‘희귀유전병’ 이기 때문에, 실제로 매우 적은 수의 사람들이 지니는 질병이다. 그렇다면 특정 유전자의 특정 돌연변이를 사람들 중 얼마나 지니고 있는지 어떻게 파악할 수 있을까? 첫 번째 방법은, database 를 이용하는 것이다. 앞서 언급된 바와 같이, 1000genomse browser를 이용하면, 기존에 천 여명을 대상으로 연구된 genome정보들 중, 인종별 allele frequency가 포함되어있다. 따라서 특정 SNP ID 에 대한 대립형질에 대한 빈도를 알 수 있다. 그러나 전세계 모든 민족에 대한 allele frequency는 포함되지 않은 단점이 있다.

그래서 파키스탄인들의 집단 내에서는 어떠한 빈도로 돌연변이가 존재하는지 직접 allele frequency를 측정해보기로 하였다. 중요한 것은, ‘아무런 특이적인 질병을 가지지 않는 정상인’ 파키스탄인 217명의 혈액으로부터 DNA를 추출한다는 것이다. 이들은 무작위로 혈액을 기증하기 위해 모인 사람들이며, 혈육관계가 있는 부모자식관계는 제외하고 형질 빈도를 파악하기로 하였다.

실험방법은 다음과 같다. Bioinformatic analysis로 추려진 후보유전자 염기서열 내에서 SNP 위치가 포함되도록 PCR primer를 제작한다(Table 1). PCR을 이용하여 특정부위를 증폭시킨 후, sanger sequencing으로 genotyping을 한다. 일반인들의 돌연변이 SNP 빈도를 파악한다. Minor allele의 비율이 0%에 가까울수록 ACD를 유발하는 원인 돌연변이일 가능성이 높다. 아래에는, 실제 PCR에 이용한 후보유전자들의 primer sequence들을 나열하였다(Table 1).

Table 1. Sanger sequencing에 사용한 후보유전자들의 PCR primer sequence

Gene Name	dbSNP ID	Amino acid change	Direction	Sequence (5' → 3')	PCR product size
<i>GPNMB</i>	N A	G375V	Forward	CACATGTATCTTTTATGTTTCGTAGCA	217 bp
			Reverse	GTGGGTGACCCTAGGTGGAG	
<i>PKP3</i>	rs151201080	T459M	Forward	TTCACACCATCCCCAGAAAC	237 bp
			Reverse	CCGGTGGCGTTGTAGAAG	
<i>SLC25A22</i>	rs111277421	L150V	Forward	CCTTGTAGAGACCGGCAATG	217 bp
			Reverse	GAGGGCCAGAGCGTGTC	
<i>KIF21A</i>	rs75223821	E1201D	Forward	CATCTAAGTGCAGCAGGCATT	236 bp
			Reverse	TTGACTATCTACCAAAGTTGTATGTGC	
<i>PHRF1</i>	rs149443419	R1046Q	Forward	CACAGAGTCGTGGAGCTCAG	286 bp
			Reverse	GCTCCCTGCTCTTGTCCTT	
<i>RASSF7</i>	rs2242182	R223Q	Forward	AGCAGTGGCCGCACTTAG	237 bp
			Reverse	GAGCCCCGCAAACACT	
<i>NUB1</i>	rs113920358	R9C	Forward	CCCACAACCCTCTTTCCAAG	207 bp
			Reverse	CAAGCAGTGCTGCTGAGG	

## (6) 돌연변이로 인한 단백질 기능 변화 예측

아미노산 서열 변화로 인한 단백질의 기능변화 영향력 예측은 총 4가지 프로그램을 이용하여 수행되었다. Polyphen, PROVEAN, SIFT, mutation taster 이다. 동일한 아미노산으로 같은 조건을 주었다 하더라도 프로그램마다 결과가 조금씩 달랐기 때문에, 여러 프로그램을 실행시켜 결과를 조합하였다.

Polyphen에서는 ① Protein sequence 정보를 기입하고, ② 돌연변이 치환 위치를 숫자로 기입, ③ 치환(Substitution)되기 전,후 아미노산의 종류를 선택한 후, 프로그램을 실행하면 예측 결과가 나온다. 특정 유전자의 전체 아미노산 서열을 알아야 하며, missense mutation만 예측 가능하다. 결과는 'probably damaging', 'possibly damaging' 혹은 'benign' 으로 나오며, 결과자체가 나오지 않는 경우도 많았다.

PROVEAN(Protein Variation Effect Analyzer)과 SIFT(Sorts Intolerant from Tolerant)은 같은 홈페이지에서 구동되며, SNP뿐만 아니라 InDel에 의한 예측 변화도 알 수 있다. 이 두 프로그램은 아미노산이나 염기 서열을 알 필요 없이 '염색체 번호, 염색체 내 돌연변이 위치, 기존 아미노산, 치환 후 아미노산' 만 기입하면 database 상에서 sequence를 검색해서 예측해준다. 예를 들어, *GPNMB* 의 아미노산 변화를 예측하는 경우, '7, 23306205, G, V' 라고만 기입하고 프로그램을 돌려주면 예측 결과를 알 수 있다. 결과는 PROVEAN의 경우 'Deleterious' 와 'Neutral' 로 나오고, SIFT의 경우에는 'Damaging' 과 'Tolerated' 로 나뉘어서 나온다. Parameter의 경우, 'Human GRCh37 Ensembl 66' 을 이용하였다.

Mutation Taster 프로그램은, ① 유전자 이름을 기입하면, 특정 유전자에 해당되는 transcription version들이 나열된다. ② 그 중, 원하는 형태의 ENST number와 NM number를 확인하여 클릭하고, ③ 치환되는 염기서열의 위치와 치환 후 아미노산을 기입한다. 결과는 'Disease causing' 과

'polymorphism' 으로 나뉘어 나오며, 단순한 예측 외에도 많은 정보를 보여준다. Substitution과 InDel 모두에서 사용 가능하다.

## (7) Vector 디자인

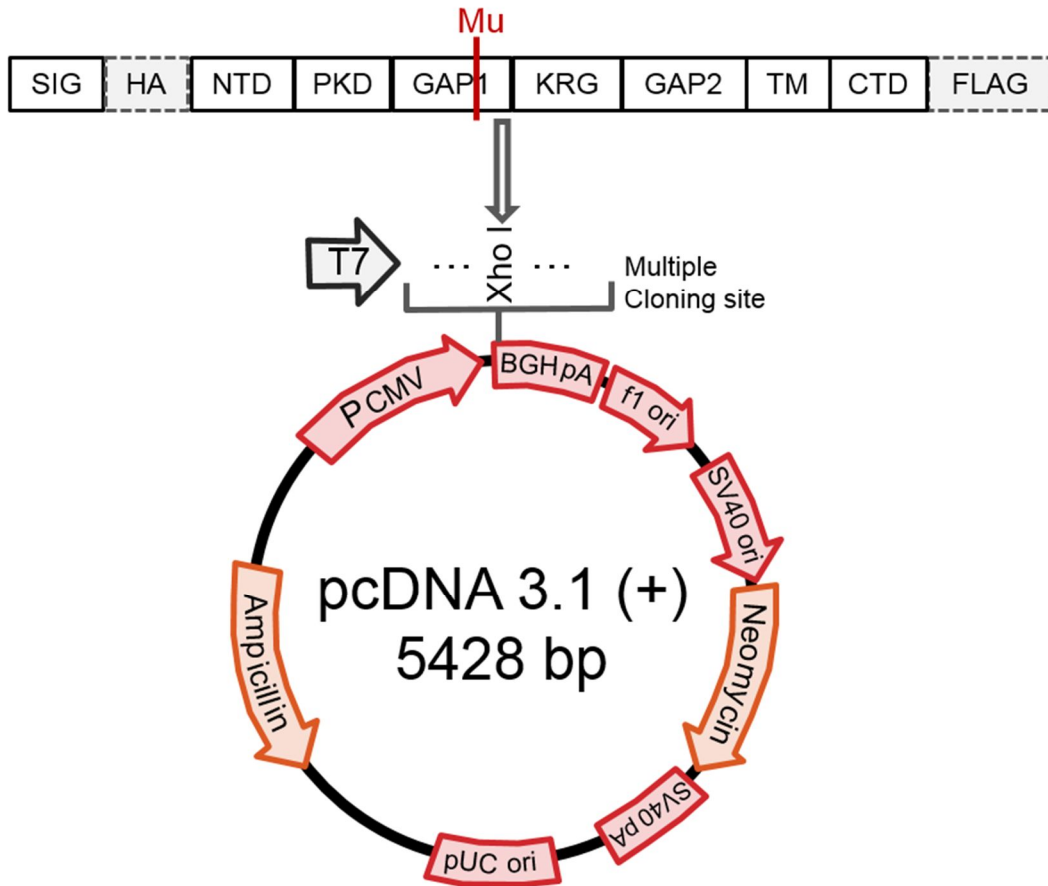
ACD를 유발하는 원인유전자가 *GPNMB*인 것으로 확정된 후, functional study를 추가로 실행하기 위해서 가장 먼저 *GPNMB* sequence가 삽입된 vector를 제작하였다. 벡터란, 재조합 DNA를 운반 및 복제하는 플라스미드 또는 박테리오파지 DNA를 의미한다. Vector로는 pcDNA3.1(+) (invitrogen)를 이용하였고, 추가 sequence 삽입 전 기존 길이는 5428bp였다(Figure 2).

vector에는 Ampicillin과 Neomycin에서 살아 남을 수(resistant) 있도록 하는 selective marker도 포함되어있다. 이 특성을 이용하여 항생제를 넣은 배지 내에서 *GPNMB*를 지니고 있는 vector만 선택적으로 살아남게 된다.

기존 vector의 multiple cloning site 내에 있는 '*XhoI*' 부위에 *GPNMB* sequence를 삽입하였다. 기존 human *GPNMB* 염기서열 양쪽에 *XhoI* sequence(5' -CTCGAG-3')를 넣었고, epitope(항원결정기) tag로 쓰이는 tagging protein의 염기서열도 함께 넣었다. HA는 앞쪽 *XhoI* 염기서열과, *GPNMB*의 가장 앞쪽, SIG(signal peptide)에 해당하는 염기서열 뒤에 넣었다. FLAG tagging protein의 염기서열은 바로 뒤에 *GPNMB*의 stop codon, 그리고 그 뒤에 뒤쪽 *XhoI* 염기서열이 위치하도록 하였다.

이렇게 해서 벡터의 총 길이는 기존 벡터 5,428 bp, 그리고 디자인한 *GPNMB*의 삽입 염기서열 1,734 bp까지 하여 총 7,162 bp였다.

Figure 2. Vector 디자인



\* Ampicillin : 항생제 ampicillin에 저항성을 지니는 염기서열, Neomycin : 항생제 neomycin에 저항성을 지니는 염기서열, P CMV: Cytomegalovirus promoter, pA: polyadenylation(termination), ori: origin, BGH: Bovin growth hormone, T7: T7 promotor, SIG: signal peptide, NTD: N-terminal domain, PKD: polycystic kidney disease-like domain, KRG: a kringle-like domain(binding interaction에 관여), TM: transmembrane domain, CTD: C-terminal domain, HA, FLAG: tagging protein

## IV. 결과

### 1. ACD환자의 임상적, 조직학적 특징

임상적으로 ACD 진단은 파키스탄 카이버 의과대학 공동연구팀에서 수행하였다. 우선 부모와 전 세대들에게 없었던 피부색소이상(피부색소이상)이 자녀들 5명 중 3명에게 나타나는 것으로 보아 전형적인 열성유전병으로 추측된다.

환자들은 과색소(hyperpigment)와 저색소(hypopigment)를 동반하는 그물모양(reticular)의 얼룩 반점들을 지니고 있었다(Figure 3). 특히 양쪽 사지와 몸통에 반점들이 많았고, 시간이 지남에 따라 반점이 목을 타고 점차 얼굴 쪽으로 번져오는 특징을 보였다. 이 병은 다른 전신증상(systemic symptom)은 나타나지 않는데, 본 연구의 대상 환자들 또한 동일한 임상양상을 보였다.

환자들 모두 태어났을 때부터 이런 피부색소이상증세를 보였던 것은 아니었다. 태어난 지 10년이 지나고 사춘기 때 즈음부터 점차적으로 반점이 생기기 시작하였다.

사춘기 이후부터 시간의 흐름에 따라 점차적으로 반점이 생겨 나가는 점, 과색소와 저색소가 공존하는 수많은 반점들, 색소이상 외에 다른 증세는 없다는 점, 부모세대에게는 나타나지 않은 열성 유전병이라는 점들로, 이들이 가지고 있는 병은 ACD임을 진단할 수 있었다.

표현형을 보고 임상적 진단을 내리는 것 이외에, 실제 피부조직에 amyloid 침착이 있는지 확인하기 위해 조직학적 검사, Skin Biopsy(피부조직 검사)를 실시하였다. 파키스탄 공동연구팀으로부터 환자의 피부조직을 paraffin block에 담아, H&E staining을 한 슬라이드를 전달받은 후, 추가로 Congo red staining과 면역조직화학염색법(immunohistochemical stain)을 진행하였는데, 항체로는 Cytokeratin 5&6을 이용하였다.

피부조직은 크게 세 부분으로 나눌 수 있는데, 피부조직 중 가장 바깥쪽에 있는 부분을 표피(Epidermis)라고 하고, 그 아래의 땀샘, 혈관 등 구조물 등이 존재하는 중간부분을 진피(Dermis)라고 하며, 진피 아래에는 피하지방으로 이루어져있다. ACD의 경우 유두진피(Papillary dermis)라고 불리는 곳에 amyloid가 침착되는 특징이 있다. 유두진피는 표피와 진피의 경계에서 돌기 같이 나와있는 부분을 말한다. 이 부분에 이상이 있는지 피부조직검사를 통해 현미경사진으로 관찰할 수 있다.

실행한 3가지 염색법 중, Congo red stain을 통해 나온 사진을 먼저 살펴보면 (Figure 4(a)), 표피에서 가장 바깥쪽인 피부각질층(Stratum corneum)과 투명층(Stratum lucidum) 사이에 비정상적인 공간이 있는 것을 알 수 있다. 이는 수포성 손상(Bullous lesions)을 입은 상태인 것으로 추측된다. 가장 바깥부터 보라색 부분까지가 표피(epidermis) 부분인데, 보라색 알맹이로 보이는 부분이 염색된 세포의 핵으로 보이며, 이 때 세포는 각질형성세포(keratinocyte)이다. Melanin세포를 만드는 Melanocyte는 표피와 진피의 경계인 기저막 (Basement membrane)에 존재하며, 각질형성세포(Keratinocyte)는 기저막부터 표피 바깥쪽까지 존재하는데, 이 세포의 상태에 따라 층이 나뉜다. 그리고 희미하게 표피에 가까운 진피와 피하 부분에 아밀로이드의 무정형 호산구 (eosinophilic amorphous) 덩어리가 침착된 것이 관찰된다. Congo red로 염색된 것을 편광렌즈를 통해 보았기 때문에, 본래 붉은색으로 나타났어야 하는 부분들이 밝은 녹색(apple-green)으로 나타났다.

보고된 ACD환자들의 피부조직 현미경 사진을 보면, Papillary dermis 부분에 아밀로이드가 침착된 것을 관찰할 수 있었다(Qiao *et al.*, 2012). 하지만, 한 환자에 대한 조직사진도, 어떤 염색법을 사용했느냐에 따라서 침착된 부분이 명확하기도, 거의 알아보기 어렵기도 하였다. 다른 Amyloid 침착 사진들도, 사실은 명확하게 보통 피부조직들과 구분이 되는 것보다는 실제로 침착이 있는 것인지의 여부는 확인하기 어려웠다. 본 연구에서 수행한 Congo red

staining, Immunohistochemical Staining, H&E staining, 3가지 염색법을 이용하여 피부조직 샘플을 따라 모두 관찰해 보았다. 하지만 명확하게 색깔로, 혹은 모양이 확연하게 다른 부분은 나타나지 않았다.

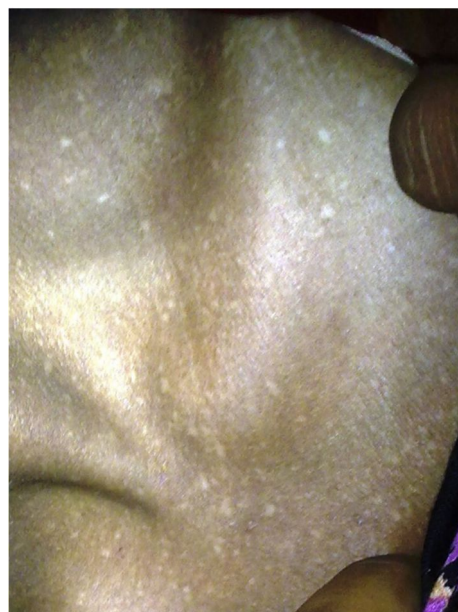
Immunohistochemical staining 사진(Figure 4(b))를 보면 확연하게 아밀로이드 침착은 나타나지 않지만, 동그랗게 들어가 있는 표피 가까이의 진피조직이, 그 아래의 진피조직들에 비해 비어있는 부분 없이 채워져 있는 것을 관찰할 수 있다. H&E staining(Figure 4(c))에서는 유두진피 내에 다른 조직에 비해 더 진한 보라색으로 염색된 알맹이들이 관찰되는 것을 볼 수 있었다. 하지만 이들은 아밀로이드의 침착으로 단정짓기에 적합하지 않다.

피부생검과 현미경사진을 통해 아밀로이드의 침착정도를 알아보려고 했지만, 실제로 관찰되는 바는 아직 없었다. 하지만 그들에게 피부색소이상이라는 것은 확실하기 때문에 이들이 지닌 피부병의 원인유전자를 찾는 과정을 그대로 수행하였다.

Figure 3. ACD 환자의 임상적 소견



(a) 등



(b) 목/ 상측 흉부

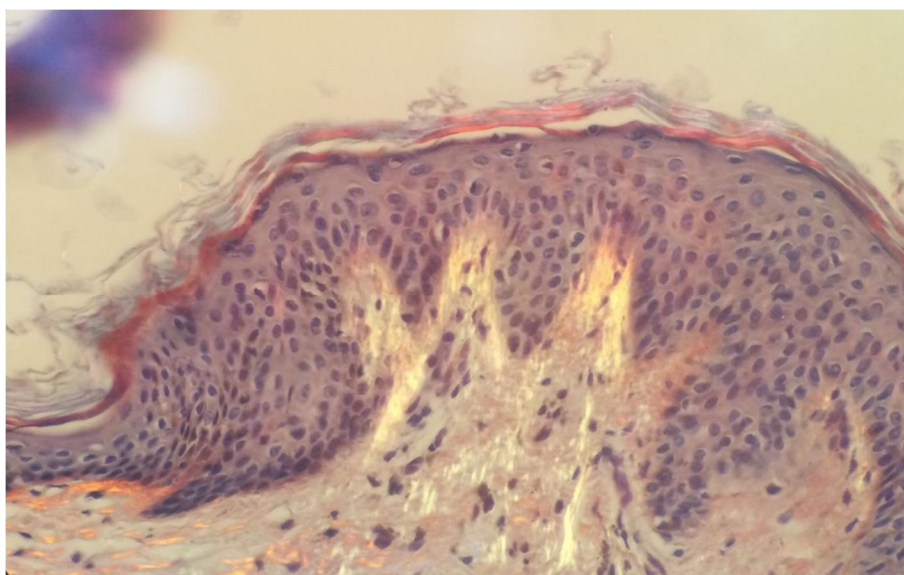


(c) 다리

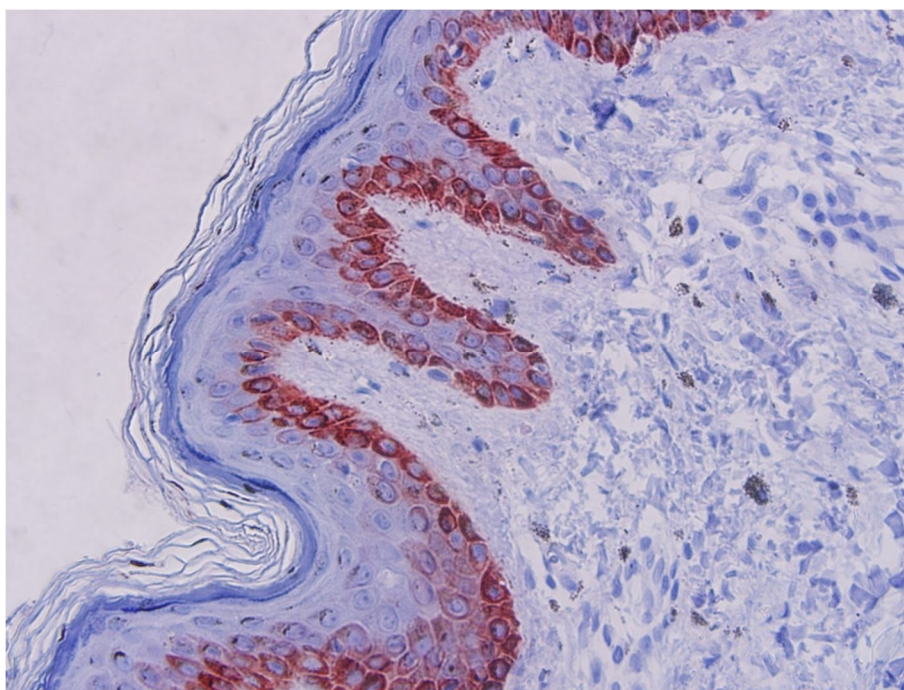


(d) 복부/몸통

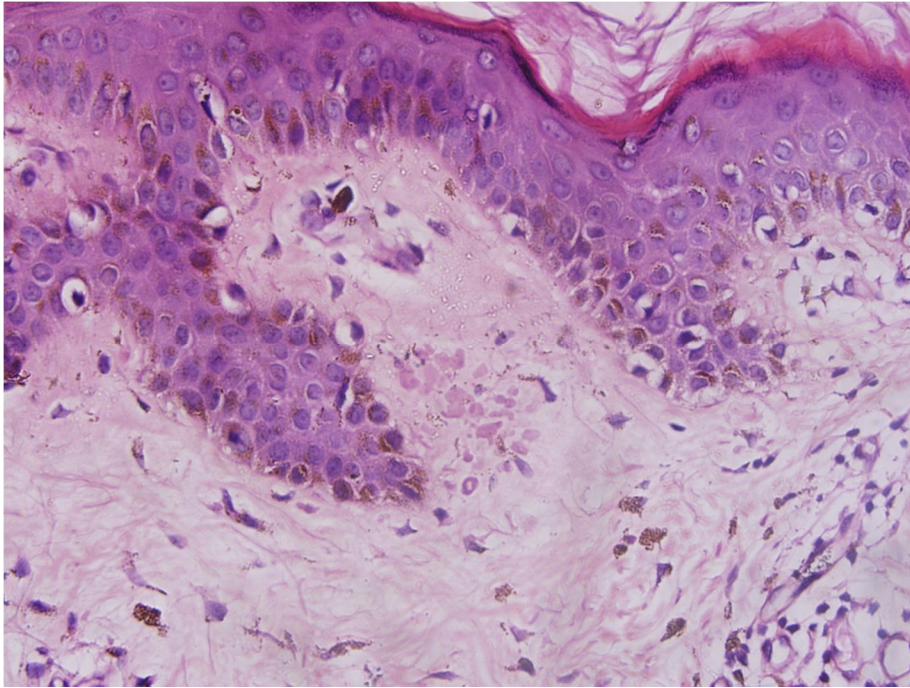
Figure 4. ACD 환자의 Skin Biopsy(피부생검) 현미경 사진



(a) Congo red staining



(b) Immunohistochemical stain (using Cytokeratin 5,6)



(c) H&E staining

## 2. 가계도 내 염기서열 분석 결과

Pedigree 내에서 정상인 부모 (IV-1, IV-2), 그리고 자녀세대에서 환자 3명 (V-2, V-3, V-5)과 정상인 자녀 2명 (V-1, V-4) 을 포함한 7명의 혈액에서 추출한 DNA를 대상으로 Next Generation Sequencing의 한 분야인 Exome Sequencing을 수행하였다(Figure 1). 약 6주에 걸쳐 NGS 분석을 마친 후, fastq 파일 형식으로 개개인의 sequencing 정보가 생성되었다.

Sequencing 완료된 염기서열 96%의 Phred score는 20이상이었으며, 총 염기 수는 9.6 Gb였다. 후보염기서열들의 평균 depth는 101이었으며, minimum Q call > 20와 minimum depth > 10를 만족하는 변이는 총 111,453개였다.

Fastq 파일은 Forward sequence, Reverse sequence로 1인당 2개씩 파일이 생성되었으며, 짝을 이루어 'SeqMan NGen' 프로그램을 이용하여 Assembly를 수행하였다. 하루에 걸쳐 assembly가 완료되면, ArrayStar program을 이용하여 각각 개인에게 조건을 주어 ACD를 유발할 수 있는 후보 원인 염기서열 변이들을 스크리닝 하였다.

부모세대(IV-1,IV-2)는 'Heterozygous' 로 다른 형질(돌연변이 형질과 정상 형질일 것으로 간주)을 1개씩 가지는 SNP, 그리고 자녀 중 환자(V-2,V-3,V-5)는 'Homozygous variation' 으로 같은 형질을 2개 가지는 SNP (variation으로 조건이 들어가 있기 때문에 정상형질을 2개 가지는 경우가 아닌, 돌연변이 형질을 2개 가지는 경우만 선택된다)을 조건으로 추려내었다.

Table 2. ACD 후보유전자들의 염기서열변이 목록

Gene Name	<i>PHRF1</i>	<i>RASSF7</i>	<i>KIF21A</i>	<i>SLC25A22</i>	<i>PKP3</i>	<i>NUB1</i>	<i>GPNMB</i>	<i>SOAT2</i>	
chromosome	11	11	12	11	11	7	7	12	
Reference/Minor allele	G/a	G/a	C/a	G/c	C/t	C/t	G/t	C/t	
dbSNP ID	rs149443419	rs2242182	rs75223821	rs111277421	rs151201080	rs113920358	N A	rs2272296	
Minor allele frequency	PJL	0.11	0.10	0.05	0.01	0.00	0.15	0.000	0.10
	CEU	0.06	0.06	0.05	0.07	0.02	0.15	-	0.16
	CHB/S	0.02	0.15	0.11	0.01	0.00	0.01		0.29
	YRI	0.00	0.00	0.01	0.04	0.00	0.00		0.37
Protein change	R1046Q	R223Q	E1201D	L150V	T459M	R9C	G375V	T254I	
Simulation of functional effect	Polyphen	benign	-	-	benign	-	benign	probably damaging	X
	PROVEAN	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Deleterious	
	SIFT	Tolerated	Tolerated	Tolerated	Tolerated	Damaging	Damaging	Damaging	
	mutation taster	disease causing	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	disease causing	

\*SNP : Single Nucleotide Polymorphism, \*CEU: Northern and Western European,

\*CHB/S: Han Chinese in Beijing/ Southern Han Chinese, \*YRI : Yoruba in Ibadan, Nigeria,

\*PJL: Punjabi in Lahore, Pakistan, \* RPP : Random Pakistani Population,

Next Generation Sequencing을 이용하여 데이터화 시킨 가계도 내 개인별 genome 정보로 DNASTAR program을 이용하여 Filtering (조건: 부모세대는 보인자 이므로 각각 다른 형질을 1개씩 가지는-heterozygous- SNP, 자녀세대 중 환자들은 같은 형질로 2개 가지는-homozygous- SNP으로 추린다)해 보았을 때, 조건에 부합되는 결과로 처음에 총 56개의 후보유전자들이 추려졌다(Figure 5).

Array Star에 준 조건만으로 추려진 56개의 후보유전자 중, 자녀 세대 중 정상인인 V-1, V-4가 자녀 세대 중 환자인 V-2,V-3,V-5와 같은 유전형질을 지니는 경우를 제외시킨다. 자녀세대의 정상인은 프로그램 중에 조건을 주지는 않았지만, 자녀세대의 환자와 다른 유전형질을 지니기 때문에 표현형도 환자가 아닌 정상으로 나타난 것이라고 간주하며, 환자들의 유전형질이 겹치는 부분들은 제외시킨다. 이러한 조건으로 추려냈을 때, 42개의 후보유전자가 남았다.

ACD family 외에, in-house 파키스탄 정상인 엑솜 정보와도 비교하여 42개의 후보 염기서열변이들 중 이들 25명의 정상인들 중 최소 1인이 동형접합 유전자형을 보일 경우 후보에서 제외하였다.

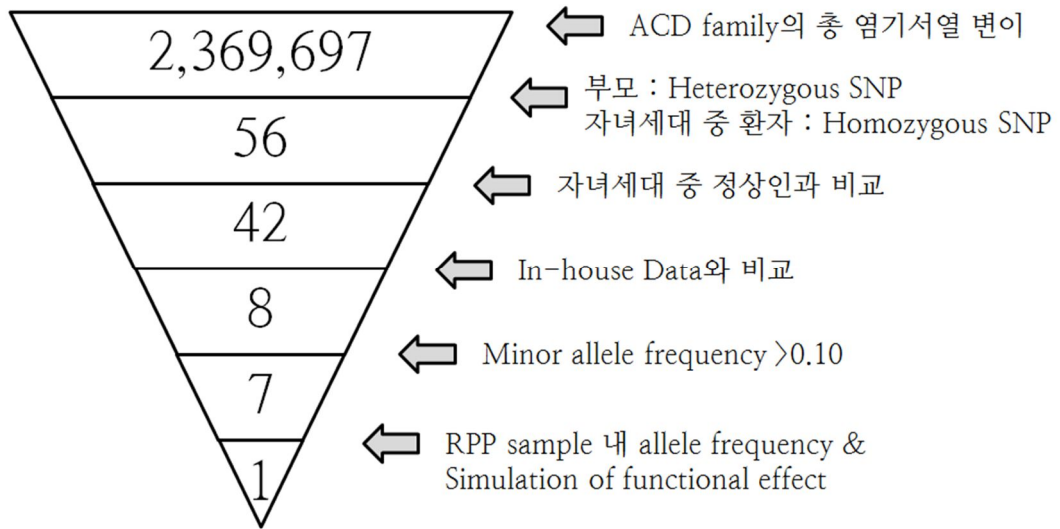
결론적으로 ACD 가계도 내에서 i) 부모세대는 이종접합 유전형질(heterozygous variation)이면서 ii) 자녀세대의 환자는 homozygous minor allele을 가지는 경우, iii) 자녀세대의 정상인과 ACD 가계도 외의 정상인들이 자녀세대의 환자와 다른(heterozygous variation 혹은 homozygous reference allele) 유전형질을 가지는 경우만 추려졌을 때, 8개의 후보유전자만이 남았다(Table 2).

8개 후보유전자를 대상으로 처음 진행했던 것은 바로 NGS 결과로 얻은 genotype에 오류가 없는지 확인하는 것이었다. Genotyping을 위해 Sanger sequencing을 이용하였으며, 후보유전자들의 SNP위치를 찾아 돌연변이 여부를 확인할 수 있는 Primer set 8개를 만들어 PCR을 진행해서 sequencing을

수행하였다(Table 1). Primer 길이는 15~27 bp(base pairs)였으며 평균 20bp 정도였다. PCR product의 길이는 200~400bp가 되도록 디자인하였다. Sanger sequencing 결과 8개 후보유전자의 NGS 결과와 동일하게 나왔다.

다음과정으로 database를 이용하여 각 후보 돌연변이들에 대한 minor allele frequency를 검색하였다. dbSNP ID가 없는 *GPNMB* 를 제외하고는, 나머지 중 6개 후보유전자의 PJI 인종(남아시아인)에서의 대립유전자빈도는 0.00~0.15였고, 평균 대립유전자빈도는 0.01~0.08였다(Table 2). 하지만 *SOAT2* 의 경우 PJI 인종에서의 대립 유전자빈도는 0.10이었지만, 백인, 흑인, 아시아인에서의 평균 대립유전자빈도가 0.23으로 높았기 때문에 후보유전자에서 제외시켰다. 다른 6개의 후보유전자들도 대립유전자빈도가 최고 0.15인 인종이 있지만, 그 수가 많지 않으므로 후보유전자 7가지를 그대로 다음 과정을 진행하였다.

Figure 5. Candidate gene filtering 과정



### 3. 파키스탄 인종 내 유전자형 분석

가계도 내 혹은 in-house 데이터들과 비교할 때, ACD 환자들은 유전형질이 겹치지 않는 후보유전자로 추려내고, database상에서 특정 유전자의 특정 위치의 대립유전자빈도를 검색하였다. 여기에 추가로, 실제로 본 연구실에서 보유한 217명의 파키스탄 정상인들의 DNA로 실제 대립유전자 빈도를 구해보았다.

대립유전자 빈도(allele frequency)란 한 개체군 내에서 하나의 특정한 대립유전자의 비율이다. 여기에서 217명의 파키스탄 정상인들은 무작위로 선출된, 서로 혈육관계가 없는 사람들이다. 이들의 혈액으로부터 DNA를 추출해내고, 특정 SNP들의 염기서열을 알아낼 수 있도록 위에서 가계도 내 개인의 NGS결과를 확인했을 때 썼던 유전자 별 Primer를 이용하여, PCR과 Sanger sequencing을 진행하였다.

처음엔 8개의 후보유전자에 대하여 122명의 DNA로 Sequencing을 수행하였다. *GPNMB*에서는 모두가 우성동형접합성(Homozygous Reference alleles)을 보였다(MAF: 0.00). *PKP3*에서는 2명(118명 sequencing 성공, MAF: 0.01), *SLC25A22*에서는 2명이(121명 sequencing 성공, MAF: 0.01), *KIF21A*에서는 11명(121명 sequencing 성공, MAF: 0.05)이 이종접합성(Heterozygous alleles)을 보였다. *PHRF1*, *RASSF7*, *NUB1*은 MAF가 각각 0.13, 0.11, 0.16이었지만, 결정적으로 ACD 환들과 같은 유전형질인 열성동형접합성(Homozygous mutant alleles)을 보이는 사람이 각각 2,1,2명이 있었다. 가계도 내 정상인들 혹은 in-house 데이터에서와 마찬가지로 이들에게는 ACD와 같은 유전형질인 사람이 단 한 명이라도 있으면 안 되기 때문에, 이 3가지의 후보 유전자들도 제외시킬 수 있었다.

남은 후보유전자 4개를 대상으로 95명의 파키스탄 정상인들의 염기서열을 추가로 sequencing하였다. *GPNMB*에서는 단 한 명에게서도 Minor allele이

발견되지 않았다(MAF: 0.00). *PHRF1*, *RASSF7*, *NUB1*에서는 열성동형접합성 유전형질을 가진 사람은 단 한 명도 나오지 않았지만, 이종접합성을 가진 사람들이 많이 나왔다. Minor allele frequency는 각각 0.01, 0.02, 0.05로 나왔다.

후보유전자로 가지고 있던 7가지 유전자에서, 앞서 언급된 3가지의 경우 열성동형접합성 정상인이 나왔기 때문에 후보에서 삭제했으며, 나머지 4가지에서는 MAF가 0.05 이하로 나왔지만 특히 *GPNMB* 유전자는 MAF가 0.00이 나왔기 때문에 가장 후보유전자로서 유력해졌다.

#### 4. 돌연변이로 인한 단백질 기능 변화 예측

후보유전자 7개 모두에 대해서, 아미노산에 변화가 생겼을 때, 실제로 얼마나 큰 영향을 끼칠 확률이 있는지 프로그램들을 이용하여 예측해 보았다.

어느 염색체의 어떤 위치에서, 어떤 유전자의 몇 번째 아미노산이 어떤 아미노산으로 바뀌는지에 대한 정보는 NGS 결과를 assembly, assignment, filtering하여 얻은 정보들에 모두 나타나 있었다(Table 2). 바뀐 후 아미노산이 기존 아미노산과 비슷한 성질을 가지는지, 구조적으로는 얼마나 다른지에 따라 실질적으로 미치는 영향의 정도도 달라진다. 아미노산이 바뀐다고 하여 무조건 치명적인 것은 아니고, 크기나 성질이 비슷한 아미노산으로 바뀔 경우에는 pathogenicity가 없다.

앞서 연구방법에서 언급하였던 4가지 프로그램을 이용하여 후보유전자들의 아미노산 변화로 인한 치명성을 예측하였는데, 그 결과는 모두 상이하였다(Table 2). 같은 후보유전자의 돌연변이라 할지라도 프로그램마다 결과가 모두 달랐다. 후보유전자 7가지중 6개 유전자들이 0개 혹은 1개 프로그램에서

만 ‘Damaging’ 할 것으로 예측되는 가운데, *GPNMB* 유전자만 일관성있게 pathogenicity가 있을 것으로 시뮬레이션 되었다. Polyphen, PROVEAN, SIFT, Mutation Taster 4가지 프로그램에서 모두 ‘probably damaging’, ‘Deleterious’, ‘Damaging’, ‘disease causing’ 이라는 결과를 얻었다. *GPNMB* p.G375V 돌연변이는 단백질의 기능에 치명적인 영향을 줄 것으로 예측 된 것이다. 특히, PROVEAN과 Polyphen 프로그램에서 유일하게 negative한 결과를 얻었다.

## 5. 플라스미드 내 돌연변이생성(Mutagenesis)

기존 pcDNA 3.1(+) 벡터에 *GPNMB* 유전자를 추가로 삽입했었다. 진행하게 될 functional study에서는 WT(wild type)의 *GPNMB* sequence를 지니고 있을 때와 ACD 가계도에서 결과로 나온 돌연변이를 지니는 *GPNMB* 염기서열을 지니고 있을 때 어떤 기능적 차이가 발생하는지를 규명하고자 하였다. 그렇기 때문에, G375V(c.1124G>T)의 형질 변환이 일어난 돌연변이 *GPNMB* 염기서열을 지닌 벡터의 제작이 선행되어야 했다. Wild-type *GPNMB* 와 돌연변이 *GPNMB* 염기서열의 차이는 단 1개의 염기 차이이기 때문에, 돌연변이 염기서열을 삽입하기 위해 Mutagenesis(돌연변이 유발) 실험을 진행하였다.

Figure 6(b)의 (iii)에서, 우측의 10kb size marker를 토대로 가장 밝고 굵은 3kb부터 6kb, 8kb, 10kb까지 표시하였다. 가장 아래에서 가장 밝고 굵은 밴드가 Major band이며, 이 밴드에는 ‘Supercoiled DNA’가 위치한다. 이 밴드는 4kb와 5kb 사이에 위치한다. 기존에 디자인된 벡터의 염기서열길이 약 7.1kb인 것을 감안했을 때, 그보다 더 아래에 밴드가 위치한 것을 알

수 있다. 그리고 그 위에 얇고 희미한 밴드가 관찰되는데, 이 밴드는 size marker와 비교했을 때, 6kb와 8kb 사이에 위치한다. 기존 벡터의 염기서열 길이에 해당하는 약 7.1kb에 위치하며, 이 밴드가 linear DNA임을 알 수 있다. 그 위에 또 하나의 희미한 밴드가 있는데, 무려 10kb보다도 위쪽에 위치하였다. 이 밴드는 'Nicked circle DNA'임을 추측할 수 있다.

그리고 Mutagenesis, 인위적인 염기서열 변환이 잘 이루어졌는지 확인하였다(Figure 7). *GPNMB*의 돌연변이 부위를 확인할 수 있는 PCR primer를 이용하여, PCR 후, sanger sequencing을 수행하였다. 염기서열상에서 *GPNMB*의 1124번째 G 염기가 T로 바뀌는 점 돌연변이가 생성되도록 디자인하였으며, 그 결과로 375번째 아미노산 G(Glycine)이 V(valine)으로 코딩되도록 디자인 하였다. Sequencing 결과, *GPNMB* sequence 내의 1124번째 G염기가 T 염기로 바뀐 것을 확인했으며, 이 부분을 제외한 모든 염기서열은 동일한 것까지 확인되었다.

## 6. 유전자 발현 실험

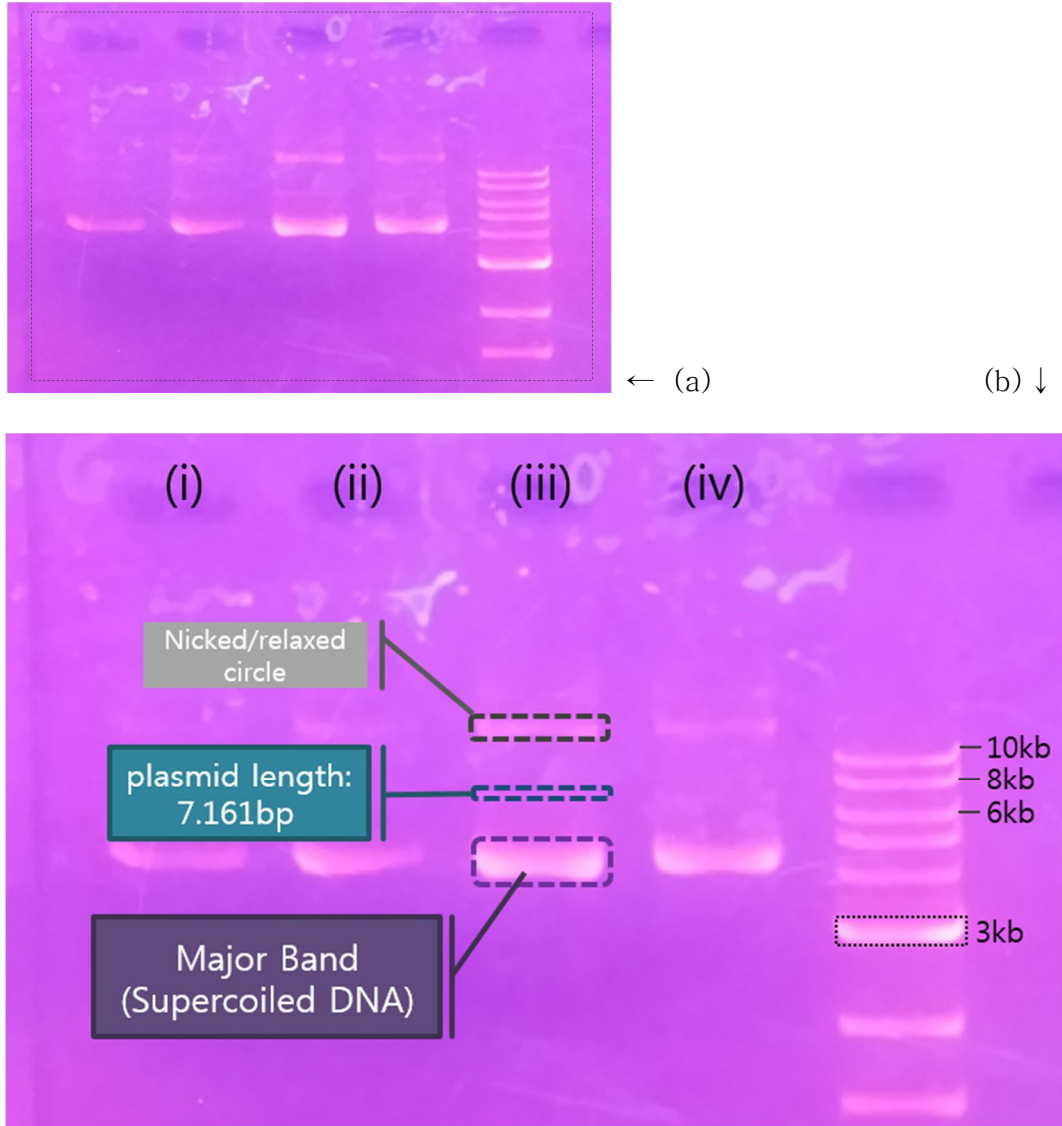
*GPNMB*유전자에서 발견한 돌연변이의 기능변이 초래여부를 확인하기 위해 먼저 *GPNMB*<sup>WT</sup>와 *GPNMB*<sup>Mut</sup>를 각각 HeLa Cell에 transfection시켰다. 본 실험은 (주)아모레퍼시픽 연구팀과 공동으로 진행하였다.

HeLa Cell line에 transfection시켜, antibody FLAG를 이용하여 western blotting을 수행하였다. Figure 8에서 EndoH (Endoglycosidase H)란, Glycan을 제거시켜 순수한 protein만 관찰할 수 있게끔 하는 효소이다. Asparagine-linked mannose rich oligosaccharides를 자르는 역할을 하며, deglycosylate glycoproteins를 찾는 목적으로 주로 이용한다. Glycosylation

은 진핵세포의 소포체에서 일어나는 대표적인 번역 후 과정(post translational modification) 중의 하나이다. 그렇기 때문에 EndoH 처리 전후 SDS-PAGE 젤 상에서의 이동 양상의 변화를 분석함으로써 GPNMB가 Golgi 혹은 ER에 위치하는지 여부를 파악할 수 있다. 하지만 본 실험에서는 EndoH 처리 후, GPNMB<sup>WT</sup>와 GPNMB<sup>Mut</sup> 벡터에서 나온 단백질의 이동양상이 다르지 않기 때문에, ER 혹은 Golgi에서 위치하는지에 대한 여부는 논의가 불가하다.

EndoH가 처리되지 않은 기존 벡터에 대한 밴드 경향을 보면, 우선 Mock(Negative Control)에서는 예측한 바대로 단백질이 관찰되지 않았다. GPNMB<sup>WT</sup>과 GPNMB<sup>Mut</sup>의 밴드패턴을 서로 비교해보면, 우선 GPNMB<sup>Mut</sup> 단백질의 경우, GPNMB<sup>WT</sup>에 비해 밴드의 intensity가 매우 약한 것을 알 수 있다. 여기서, GPNMB<sup>Mut</sup> 염기서열을 지니는 경우에 만들어지는 단백질의 Stability에 문제가 있음을 알 수 있다. 두 번째로는, GPNMB<sup>WT</sup>와 GPNMB<sup>Mut</sup>에서 밴드의 수가 차이 나는 것을 알 수 있다. 여기에서는, GPNMB<sup>WT</sup> 단백질과 GPNMB<sup>Mut</sup> 단백질의 glycosylation이 차이가 남을 알 수 있다. 이 두 가지 근거로, GPNMB에 c.1124G>T 돌연변이가 생길 경우에 만들어지는 단백질 GPNMB는 ① Stability가 낮으며, ② 기존 단백질과 glycosylation의 차이가 있음을 알 수 있었다. 따라서, GPNMB p.G375V 돌연변이로 인해 단백질의 구조와 안정성에 문제가 생긴다는 것이 입증되었다.

Figure 6. 돌연변이 생성 결과 전기영동 사진



Mutagenesis 후의 벡터를 Transformation 시켜 생성된 대장균으로부터 추출한 플라스미드의 전기영동 결과

Figure 7. WT/Mut 플라스미드 염기서열 분석 비교

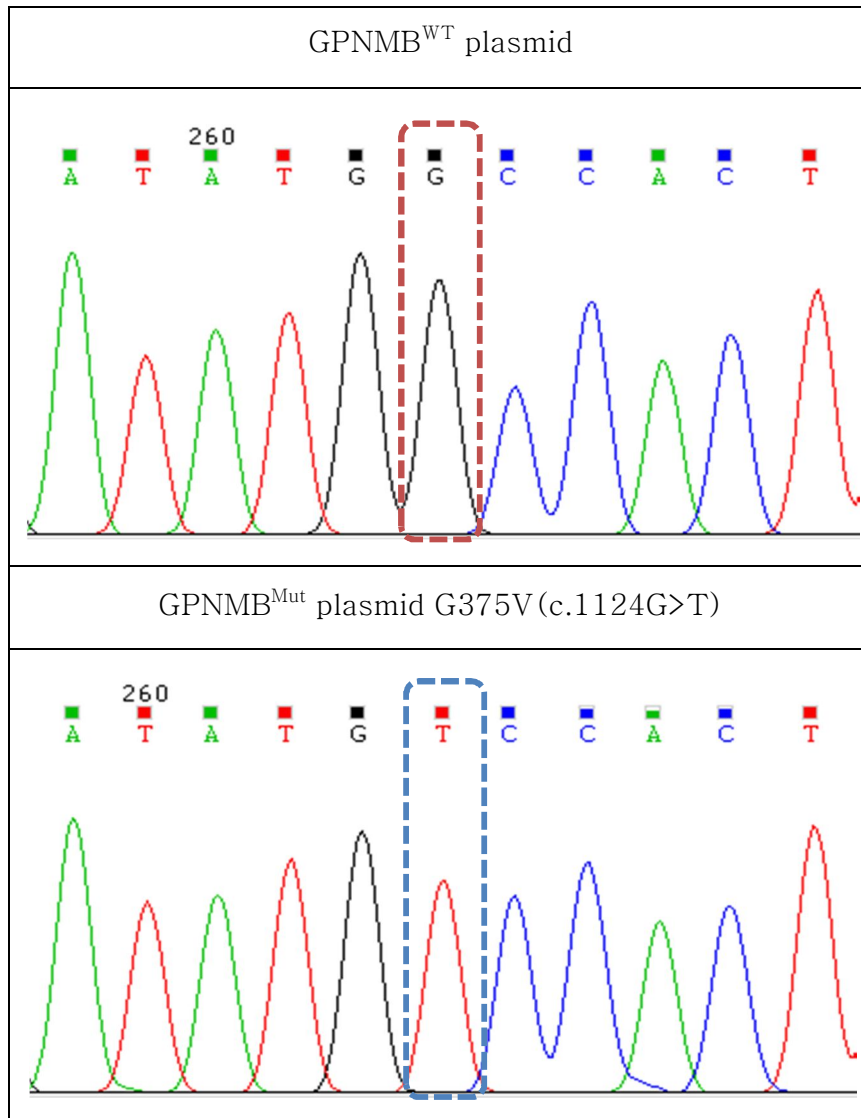
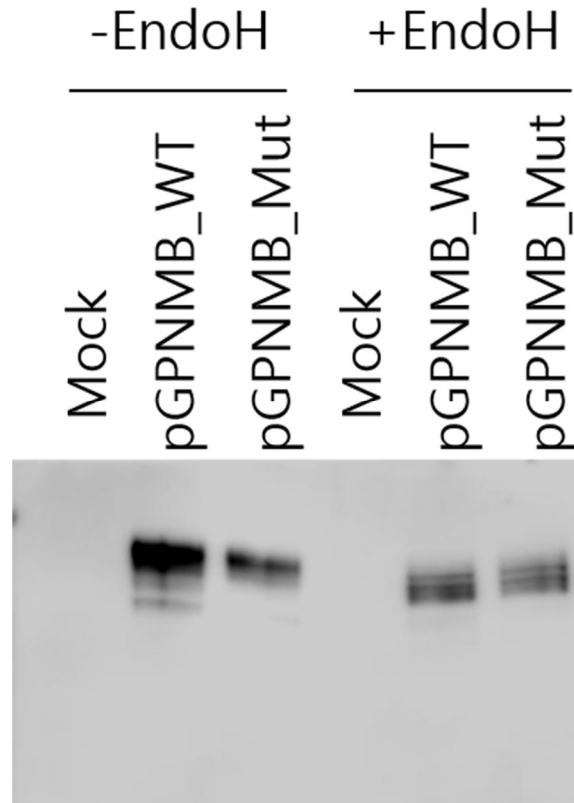


Figure 8. *GPNMB* 발현 실험



\* pGPNMB: protein GPNMB, WT: Wild type, Mut: Mutant (GPNMB p.G375V), Endo H: Endoglycosidase H

\* Endo H 처리 후, WT과 Mut의 단백질 양상이 크게 다르지 않은 것을 보아 glycosylation 전의 순수한 단백질에는 크게 차이가 없다. 반면 Endo H 처리 전, WT과 Mut 단백질의 밴드 수가 다른 것으로, glycosylation pattern에 차이가 있음을 알 수 있고, 밴드 두께로 발현량의 차이가 있음을 알 수 있다.

## V. 결론 및 논의

본 연구에서는 파키스탄에서 수집한 동계교배 가계도를 활용하여 열성 유전병인 ACD(Amyloidosis cutis dyschromica)의 원인유전자를 찾는 연구를 수행하였다. 정상인 부모에게서 태어난 자녀들 5명 중 3명이 피부색소이상증을 보이는, 전형적인 열성유전패턴을 보이는 가계도를 대상으로, 부모와 자녀 5명 모두의 혈액을 이용하여 DNA를 추출하였다. 이 DNA를 이용하여 Exome Sequencing을 수행하였고, Bioinformatics analysis 를 이용하여 후보유전자를 추려내었다.

부모에게는 Heterozygous로 존재하고, 자녀들 중 ACD 표현형을 지니는 환자들은 Homozygous로 지니는 SNP들을 스크리닝 하였고, 1차적으로 56개의 후보 염기서열변이들을 발견하였다. 추가분석을 통해 다른 정상인들에게서 발견되는 염기서열변이들은 제외함으로써 7개의 후보 염기서열변이들을 발견하였다.

7개 후보 염기서열변이들의 빈도를 파키스탄 정상인 217명을 대상으로 측정하였을 때 GPNMB p.G375V의 경우 유일하게 그 빈도가 0인 것으로 나타났다. 온라인 프로그램을 이용하여 단백질 변화에 의한 치명도 시뮬레이션에서, 아미노산 변화가 일어났을 때 치명적인 영향이 있을 것으로 모든 프로그램에서 결과가 나온 것 또한 GPNMB가 유일하였다.

이러한 시뮬레이션 결과는 GPNMB가 ACD의 원인유전자일 가능성이 높다는 것을 의미한다. 그러나 G375V, 사실 아미노산 G(Glycine)과 V(Valine)은 모두 비극성으로 성질이 같은 아미노산이다. 치명적인 결과를 초래하는 이유는 글라이신(G; Glycine) 아미노산의 성질을 고려하면 다음과 같은 가설을 설정할 수 있다.

‘이(Glycine)는 모든 아미노산 중에서 가장 작고 기본적인이다. 이런 성질 때문에 글라이신은 다른 아미노산이 쉽게 들어갈 수 없는 작은 공간을 채울 수 있으며, 이런 성질 때문에 글라이신이 다른 아미노산으로 바뀔 경우 다른 아미노산이 공간을 제대로 채우지 못해 단백질의 구조가 바뀌기 쉽다.’

Glycine과 Valine은 비록 성질은 같지만, 유난히 작은 Glycine의 성질 때문에, 그 아미노산만이 채울 수 있던 자리에 Valine이라는 비교적 큰 아미노산이 들어가게 되면서 단백질의 형태형성에 비정상적인 영향을 주게 된 것으로 예측된다. 실제로 분자량도 Glycine 75.07, Valine 117.15로 약 1.5배 차이 나는 것을 알 수 있다.

이런 근거들을 기반으로, 유전학적인 분석을 통한 ACD의 후보유전자로는 *GPNMB*가 매우 유력하다는 결론에 도달하였다. 그렇다면, 실제로 *GPNMB*란 어떤 유전자이며, 생체 내에서 어떤 부분을 담당하고 관여하고 있는지 더 조사해야만 하였다. 뒤에서 자세히 언급하겠지만, *GPNMB*는 세포막 경계에 있는 당단백질로, melanin색소와 연관이 있다고 여러 차례 보고가 되었다. 하지만, 어떤 pathway에 관여하는지 구체적으로 보고된 바는 없었다. *GPNMB*가 피부색소에 직접적으로 영향을 주는 ‘melanin’ 색소에 연관성이 있다는 것을 알고, 그렇다면 ‘melanin’ 과 ‘melanosome’ 은 어떻게 만들어지고 어디에 위치하는지 좀 더 알아볼 필요가 있었다. 그렇다면 이제, *GPNMB*과 melanosome에 대하여 자세히 알아보려고 한다.

## (1) *GPNMB* 유전자

유전자는 ‘glycoprotein (transmembrane) non-metastatic melanoma protein b’이며, 고도로 당화(glycosylation) 되어있는 type I transmembrane protein 이다. 염색체 7번 p15에 위치하며, 560개의 아미노산을 코딩 (encoding)한다. *GPNMB* 유전자는 8개의 domain을 가지고 있는데, (i) SIG; a signal peptide, *GPNMB*가 분비 경로로 들어갈 수 있도록 한다, (ii) NTD; N-terminal domain, (iii) PKD; a polycystic kidney disease-like domain, 단백질의 접힘(folding)에 관여, (iv) GAP1, (v) KRG; a kringle-like domain, 주로 조절을 위한 다른 단백질들과의 상호작용에 관여, (vi) GAP2, (vii) TM; a transmembrane domain, (viii) CTD : C-terminal domain 으로 이루어진다(Hoashi *et al.*, 2010).

*GPNMB*는 세포막 경계에 있는 당단백질 (membrane-bound glycoprotein) 으로서, 특히 각질형성세포 (keratinocyte)에서 많이 발현된다 (Tomihari *et al.*, 2009). NTD(N-terminal domain)내에 있는 RGD motif는 접착 능력(adhesion capacity)을 지닌다(Tomihari *et al.*, 2009). 세포 내 (intracellular) *GPNMB* 단백질은 우선적으로 멜라닌 소체(melanosome)에 존재하며, 멜라닌 소체가 없는 세포의 경우엔 용해소체(lysosome)에 존재한다(Tomihari *et al.*, 2009).

*GPNMB*는 멜라닌 소체 구성단백질로 알려져 있는 Pmel17 단백질과 높은 상동관계 (homology)를 보인다. Pmel17은 초기 멜라닌 소체의 성숙에 매우 중요한 역할을 하는데, 멜라닌 소체의 초기단계(stage I, II)에 많이 존재한다. Stage I 시기에 소낭(vesicle) 안으로 들어가 루멘 내 섬유줄무늬 (intraluminal fibrillary stration, amyloid fibrillar matrix)을 형성하게 되는데, Stage II로 진행되었음을 의미한다. 이에 반해, *GPNMB*는 후기 멜라닌 소체 (stage III, IV)에 많이 존재하는데, 아마도 멜라닌 형성세포(melanocyte)

로부터 케라틴 합성세포(keratinocyte)로의 멜라닌 소체 수송(transport)을 담당할 것으로 추측된다(Hoashi *et al.*, 2010). 그리고 GPNMB의 안정성(stability)는 Pmel17보다 훨씬 낮다(Hoashi *et al.*, 2010).

GPNMB의 PKD domain이 심하게 'N-glycosylation'이 되면서 표적화(targeting)하는 역할을 잃게 되고(inactive, null), 유사성을 가지는 PMEL 단백질과 구분되던 endosome으로의 수송(trafficking)이 되지 않는다(Theos *et al.*, 2013). 하지만 glycosylation이 전혀 일어나지 않게 설정했을 때에는, GPNMB NTR-PKD 조각(fragment)이 *in vitro* 상에서 PMEL처럼 endosome에 들어가 amyloid-like fibril을 만들어낸다(Theos *et al.*, 2013). 하지만 PMEL의 멜라닌 소체 내 원섬유는 detergent-insoluble인 반면, GPNMB 단백질은 Golgi에서 떨어져 나온 직 후, 'site 2' protease에 의해 절단되고 soluble protein이 되기 때문에 두 단백질은 분자세포적으로 큰 차이점을 가지고 있다(Theos *et al.*, 2013). 이런 이유들과 더불어 GPNMB에는 원섬유를 만드는데 중요한 역할을 하는 RPT domain도 없고(Hoashi *et al.*, 2006), 엔도솜 안으로 들어가 절단되어 줄무늬를 만들 특정 단백질분해 위치(specific proteolytic cleavage site)도 없기 때문에 역시 아밀로이드 섬유를 만들어낼 수 없다(Berson *et al.*, 2003).

분비된 GPNMB(secreted GPNMB; sGPNMB)는 악성 흑색종(malignant melanoma)의 조직학적 표지(histological marker)로써 효과적인 것으로 예측된다. sGPNMB는 세포 외 영역(membrane protein의 구조에서 extracellular space로 나와있는 domain)이 절단(ectodomain shedding)되어 세포 외로 분비가 되고, 분해는 post-Golgi compartment에서 이루어지며, 분비와 분해는 빠르게 일어난다(Hoashi *et al.*, 2010). GPNMB의 P1 form은 ER-modified 형태이며 M form은 대부분 Golgi-modified 형태인데, sGPNMB의 경우 M form의 ectodomain 부분이다(Hoashi *et al.*, 2010).

자외선 B(UVB; ultraviolet B)를 쬐어줄 경우, 멜라닌형성세포 (melanocyte)내의 GPNMB 발현량은 증가한다(Zhang *et al.*, 2012). GPNMB-siRNA를 형질주입 (transfection)시킬 경우, GPNMB의 발현량이 현저히 줄며 동시에 멜라닌형성세포 (melanosome)의 형성 또한 적어진다 (Zhang *et al.*, 2012). 멜라닌색소를 만드는 과정에서 첫 번째, 각질형성세포 (keratinocyte)가 분비하는  $\alpha$ MSH( $\alpha$ -melanocyte-stimulating hormone)을 인위적으로 처리해 줄 경우에도 GPNMB 단백질의 발현량이 훨씬 증가한다(Tomihari *et al.*, 2009).

*GPNMB* 유전자의 돌연변이로 사람에게 발병되는 질병은 아직 보고된 것이 없지만, 쥐의 경우에는 다르다. 쥐의 *Gpnmb* 유전자의 경우, 총 574개의 아미노산으로 이루어지는데, 그 중에 150번째 R(Arginine)이 종결코돈(Stop codon)으로 바뀌었을 때, IOP (intraocular pressure;안압)이 높아졌고, 색소성녹내장 (pigmentary glaucoma)이 발병하는 것으로 알려져 있다 (Anderson *et al.*, 2002). IPD(iris pigment dispersion; 홍채색소 분산)과 ISA(iris stromal atrophy; 홍채지질 위축증)이 동반되는 색소성녹내장이 발병했는데, 이 두 가지 증상은 2가지의 melanosomal protein gene에 생기는 돌연변이로 생겼다. IPD의 경우, *GPNMB* 유전자 150번째 아미노산에 stop codon이 생기는 돌연변이를 동종접합 (homozygous)로 가지고 있어야만 나타났고, ISA의 경우 Tyrp1 (Tyrosinase-related protein 1; OMIM 115501 ; 멜라닌 형성세포의 3번째 단계에 가서 색소형성을 시작하고, 멜라닌세포의 성숙에 직접적으로 관여) 유전자에 돌연변이가 생길 때 발생된다(Anderson *et al.*, 2002).

## (2) 멜라닌 소체 및 멜라닌 형성

멜라닌 소체(melanosome)는 세포 내 막소기관 (intracellular membranous organelle) 중 하나로, 리소솜 관련 단백질(lysosome-related protein)이다. 표피 멜라닌 세포(epidermal melanocytes)와 안구 색소세포(ocular pigment cell)에서 멜라닌 색소(melanin pigments)를 합성하고 저장하는 역할을 한다(Hearing, 2005). 피부 색이나 눈동자 색과 연관이 있으며, 빛으로부터 보호 (photoprotection) 의 역할도 지닌다(Wasmeier *et al.*, 2008).

멜라닌 소체의 발달 과정은 4단계로 이루어지는데, (1) 색소 없는 초기 엔도솜, (2) 내부 줄무늬 생성 (figrillar matrix, internal striation), (3) 멜라닌 색소 형성과 발달, (4) 중합된(polymerized) 멜라닌색소로 채워지는 과정을 거친다(Seiji *et al.*, 1963). 초기 단계에는 색소 세포-특이 단백질인 PMEL의 단백질 분해 조각(proteolytic fragments)들이 많고(Raposo *et al.*, 2001), 후기 단계에는 멜라닌 색소 발달에 관여하는 효소들 (tyrosinase, TYRP1)(Theos *et al.*, 2005), 혹은 이 효소들의 수송에 관여하는 효소들(OCA2)이 많다(Raposo and Marks, 2007; Sitaram *et al.*, 2009).

Melanin색소를 만드는 Melanogenesis과정은 다음과 같이 이루어진다. (1) 외부자극(자외선)에 노출되면 각질형성세포(keratinocyte)로부터  $\alpha$ MSH ( $\alpha$ -melanocyte-stimulating hormone)가 분비된다. (2)  $\alpha$ MSH는 피부 멜라닌형성세포의 세포막에 위치하는 MC1R(melanocortin 1 receptor)을 자극시킨다. (3) 그 결과, cAMP-dependent signalling이 일어나며 MITF (microphthalmia-associated transcription factor)의 분비를 자극하게 된다(Wasmeier *et al.*, 2008). 여기서 MITF는 멜라닌 색소를 형성하는 과정의 중요한자인데, 멜라닌 형성세포의 기능과 멜라닌형성 (melanogenesis)의 'master regulator' 라고 할 수 있다. MITF는 멜라닌 형성세포의 (a) 생존,

(b) 운동성, (c) 분화와 세포 소멸 (apoptosis), (d) 멜라닌 소체 생산에 관여하는 유전자들의 발현을 조절한다 (Wasmeier *et al.*, 2008).

생체 내 색소형성에 관련된 유전자는 150여개지만, melanosome-specific protein은 12개이다. (a) tyrosinase Tyr, (b) tyrosinase-related protein 1 (*Typ1*), (c) tyrosinase-related protein 2 (*Trp2*, dopachrome tautomerase), (d) ocular albinism type 1 protein (*OAI*), (e) melanoma-associated antigen recognized by T cells (*MART-1*), (f) Pmel17/gp100, (g) vesicle amine transport protein 1 homolog (*VAT-1*), (h) oculospanin, (i) syntenin, (j) flotillin-1/2, (k) coiled-coil-helix-coiled-coil-helix domain containing 3 (*CHCHD3*), (l) glycoprotein (transmembrane) non-metastatic melanoma protein b (*GPNMB*) (Basrur *et al.*, 2003; Hoashi *et al.*, 2010). Tyr, Trp1, Trp2와 같이 멜라닌 형성과정에서 효소작용을 하는 단백질들은 그 역할이 잘 알려져 있지만(Olivares and Solano, 2009), GPNMB를 포함한 구조 단백질들의 멜라닌 소체 내 생합성에 대해서는 확실히 알려진 바가 없다(Hoashi *et al.*, 2010).

성숙한 멜라닌 소체는 멜라닌 형성세포의 말단을 통해 각질형성세포로 전달이 되는데, 이 과정이 정확히 어떻게 이루어지는지는 아직 보고된 바 없다. 가능성 있는 방법에는 (a) melanocyte와 keratinocyte의 막 융합 (membrane fusion), (b) 소낭을 이용한 수송 (vesicular transfer), (c) 식세포작용 (phagocytosis), (d) 세포포식작용 (cytophagocytosis) 등이 있다 (Wasmeier *et al.*, 2008).

기존 *GPNMB* 염기서열을 지니는 벡터와 염기서열에 변화를 준 돌연변이 벡터를 만들었고, 그 두 가지 벡터에서 만들어내는 단백질이 분명한 차이를 보임도 알게 되었다. 이제 유전학적인 측면에서 ACD 유전병의 원인유전자는 규명 되었다. 그러나 추가적인 functional study를 수행하여 *GPNMB*의 비정상적인 기능이 ACD를 유발하는 분자 기작을 규명하는 연구가 추가적으로 진행될 필요성이 있다.

*GPNMB*가 melanin 세포와 관련이 있다는 점, 그리고 melanocyte에서 keratinocyte로 melanosome이 옮겨지는 과정에서 binding에 관여한다는 점은 알려져 있다. 그러나 구체적으로 melanin세포와 어떤 연관이 있는지, 실제로 melanosome은 어떻게 세포간이동이 이루어지는지에 대해서는 정확히 연구된 바가 없다. 여기서 두 가지 가설을 세울 수 있다. 첫째, *GPNMB*에 생긴 돌연변이가 melanin색소의 생성 자체에 영향을 줄 수 있다는 점이다. 이 경우라면, 우선 세포 내 melanosome의 발현량과 stability, 그리고 localization 연구가 진행될 필요가 있다. WT의 경우와 비교했을 때, immunofluorescence study를 진행하여, 세포 내 *GPNMB*의 단백질의 위치 비교를 할 수 있다. 그리고 (Raposo *et al.*, 2001) 논문에 나와 있는 것처럼, IEM(Immunoelectron microscopy) analysis 를 이용하면 melanosome 자체의 모양도 관찰할 수 있다. 만약에 이 경우라면, 우리가 가지고 있는 MNT-1 cell 내에서 충분히 가능한 실험들이고, 지금껏 나와있는 실험들을 참고하여 확인만 들어가면 된다.

두 번째, *GPNMB*<sup>Mut</sup>가 melanosome의 이동과정에서 오류를 준다는 점이다. 이 경우에는 melanin색소의 형성까지는 문제가 없지만, 피부 안쪽 (melanocyte)에서 피부 바깥쪽(keratinocyte)로의 이동에 문제가 생겨, 멜라닌이 피부(papillary dermis)에 쌓이게 되고, 색소이상으로 보이게 된다는 것이다. 이 경우에, MNT-1 cell 내에서 이루어지는 melanin 색소과정에는 문제가 없을 것이다. 실제 melanocyte과 keratinocyte의 연결에서, 실제 피부 조직 내에서의 이동이 관건이기 때문에 단순히 1개의 세포만으로 실험을 할

수 없다. *in vivo* 상황이며, melanocyte에서 keratinocyte까지, dermis에서 epidermis까지 조직이 있어야만 한다. 사실, melanosome과 관련된 GPNMB의 역할로 보고된 것이 melanocyte과 keratinocyte 사이의 'binding'이기 때문에, 앞서 말한 melanin색소 자체에 문제를 형성한다기보다 이동에 오류를 주는 경우에 더 가능성이 높을 것으로 추측된다.

본 연구를 통해 GPNMB라는 유전자의 고장으로 인해 초래되는 사람의 유전병을 최초로 규명하였다. 비록 쥐의 경우에서, *Gpnmb*에 stop codon이 생길 경우, 눈에 녹내장이 생긴다는 보고는 있었지만, 사람의 경우에는 아직 없었다. 그리고 ACD라는 유전병의 입장에서도 원인유전자를 밝혀낸 첫 사례이다. 비록 ACD가 생명을 위협하는 치명적인 병도 아니고, 가려움증을 유발하여 일상에 지장을 주는 병도 아니지만, 원인유전자를 밝혀낸 것에는 분명한 의의가 있다. 그리고 GPNMB의 기능장애라는 원인으로 발생할 수 있는 1가지의 결과물이 보고가 된다면, 그 사이에 이루어지는 pathway 연구가 원활해질 것으로 기대된다.

미용적인 측면에서도, 많은 사람들이 미백에 대한 관심도가 높다. 단순히 melanosome을 퇴화시켜 피부를 하얗게 만드는 것을 연구하기보다, melanin 세포의 형성부터 이동이 어떻게 이루어지는지를 밝혀내게 되면 미백에 대한 연구도 새로운 측면에서 가능하게 될 것으로 기대된다. 위험성도 있지만, 그 모든 과정을 밝히게 된다면 색소 자체의 형성이나 이동을 조절하는 기술이 개발될 수도 있다. 하지만 미용적인 측면에서 아름다워 보이더라도, melanin 색소의 존재이유와 필요성이 있는 것은 분명하다. 어떻게 활용이 되든지 melanin과 피부색소의 연구에, GPNMB와 ACD의 관계 규명이 어떠한 영향을 주게 될 것 또한 분명한 일이다.

## 참고 문헌

Anderson MG, Smith RS, Hawes NL, *et al.* (2002) Mutations in genes encoding melanosomal proteins cause pigmentary glaucoma in DBA/2J mice. *Nature genetics* 30:81-5.

Bashiardes S, Veile R, Helms C, *et al.* (2005) Direct genomic selection. *Nature methods* 2:63-9.

Basrur V, Yang F, Kushimoto T, *et al.* (2003) Proteomic analysis of early melanosomes: identification of novel melanosomal proteins. *Journal of proteome research* 2:69-79.

Berson JF, Theos AC, Harper DC, *et al.* (2003) Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. *The Journal of cell biology* 161:521-33.

Bourke JF, Berth-Jones J, Burns DA (1992) Diffuse primary cutaneous amyloidosis. *The British journal of dermatology* 127:641-4.

Fernandez-Flores A (2012) Cutaneous amyloidosis: a concept review. *The American Journal of dermatopathology* 34:1-14; quiz 5-7.

Hearing VJ (2005) Biogenesis of pigment granules: a sensitive way to regulate melanocyte function. *Journal of dermatological science* 37:3-14.

Ho MS, Ho J, Tan SH (2009) Hypopigmented macular amyloidosis with or without hyperpigmentation. *Clinical and experimental dermatology* 34:e547-51.

Hoashi T, Muller J, Vieira WD, *et al.* (2006) The repeat domain of the melanosomal matrix protein PMEL17/GP100 is required for the formation of organellar fibers. *The Journal of biological chemistry* 281:21198-208.

Hoashi T, Sato S, Yamaguchi Y, *et al.* (2010) Glycoprotein nonmetastatic melanoma protein b, a melanocytic cell marker, is a melanosome-specific and proteolytically released protein. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 24:1616-29.

Madarasingha NP, Satgurunathan K, De Silva MV (2010) A rare type of primary cutaneous amyloidosis: amyloidosis cutis dyschromica. *International journal of dermatology* 49:1416-8.

Mahon C, Oliver F, Purvis D, *et al.* (2015) Amyloidosis cutis dyschromica in two siblings and review of the epidemiology, clinical features and management in 48 cases. *The Australasian journal of dermatology*.

Meder B, Haas J, Keller A, *et al.* (2011) Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. *Circulation Cardiovascular genetics* 4:110-22.

Mertes F, Elsharawy A, Sauer S, *et al.* (2011) Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in functional genomics* 10:374-86.

Morishima T (1970) A clinical variety of localized cutaneous amyloidosis characterized by dyschromia (amyloidosis cutis dyschromica). *The Japanese journal of dermatology* 80:43-52.

Ogino A, Tanaka S (1977) Poikiloderma-like cutaneous amyloidosis. Report of a case and review of the literature. *Dermatologica* 155:301-9.

Olivares C, Solano F (2009) New insights into the active site structure and catalytic mechanism of tyrosinase and its related proteins. *Pigment cell & melanoma research* 22:750-60.

Ozcan A, Senol M, Aydin NE, *et al.* (2005) Amyloidosis cutis dyschromica: a case treated with acitretin. *The Journal of dermatology* 32:474-7.

Qiao J, Fang H, Yao H (2012) Amyloidosis cutis dyschromica. *Orphanet journal of rare diseases* 7:95.

Raposo G, Marks MS (2007) Melanosomes [mdash] dark organelles enlighten endosomal membrane transport. *Nat Rev Mol Cell Biol* 8:786-97.

Raposo G, Tenza D, Murphy DM, *et al.* (2001) Distinct protein sorting and localization to premelanosomes, melanosomes, and lysosomes in pigmented melanocytic cells. *The Journal of cell biology* 152:809-24.

Schreml S, Szeimies RM, Vogt T, *et al.* (2010) Cutaneous amyloidoses and systemic amyloidoses with cutaneous involvement. *European journal of dermatology : EJD* 20:152-60.

Seiji M, Fitzpatrick TB, Simpson RT, *et al.* (1963) Chemical composition and terminology of specialized organelles (melanosomes and melanin granules) in mammalian melanocytes. *Nature* 197:1082-4.

Sitaram A, Piccirillo R, Palmisano I, *et al.* (2009) Localization to mature melanosomes by virtue of cytoplasmic dileucine motifs is required for human OCA2 function. *Molecular biology of the cell* 20:1464-77.

Theos AC, Tenza D, Martina JA, *et al.* (2005) Functions of adaptor protein (AP)-3 and AP-1 in tyrosinase sorting from endosomes to melanosomes. *Molecular biology of the cell* 16:5356-72.

Theos AC, Watt B, Harper DC, *et al.* (2013) The PKD domain distinguishes the trafficking and amyloidogenic properties of the pigment cell protein PMEL and its homologue GPNMB. *Pigment cell & melanoma research* 26:470-86.

Tomihari M, Hwang SH, Chung JS, *et al.* (2009) Gpnmb is a melanosome-associated glycoprotein that contributes to melanocyte/keratinocyte adhesion in a RGD-dependent fashion. *Experimental dermatology* 18:586-95.

Turner EH, Ng SB, Nickerson DA, *et al.* (2009) Methods for genomic partitioning. *Annual review of genomics and human genetics* 10:263-84.

Wasmeier C, Hume AN, Bolasco G, *et al.* (2008) Melanosomes at a glance. *Journal of cell science* 121:3995-9.

Zhang P, Liu W, Zhu C, *et al.* (2012) Silencing of GPNMB by siRNA inhibits the formation of melanosomes in melanocytes in a MITF-independent fashion. *PloS one* 7:e42955.

## ABSTRACT

### Exome–sequencing to identify genetic cause of Amyloidosis Cutis Dyschromica in a Pakistani inbred family

Jeena Kim

Department of Biology

Graduate school of

Sungshin Women' s University

Amyloidosis cutis dyschromica (ACD) is a rare variant form of cutaneous amyloidosis that is characterized by the presence of widely distributed hyper–pigmented and hypo–pigmented macules on the skin. This disorder often clusters in families and it has been suggested that genetic factors might be involved. However, no genes have yet been identified associated with this disorder. In an effort to identify the genetic causes of ACD, we have identified a consanguineous Pakistani family with multiple cases of ACD that displays a recessive mode of inheritance. We performed whole–exome sequencing (WES) in members of this family followed by bioinformatic and *in silico* analyses, and revealed that all affected individuals were homozygous for a novel missense mutation,

p.Gly375Val (c.1124G>T), in the glycoprotein nonmetastatic melanoma protein b (*GPNMB*) gene. This variant is not present in 217 Pakistani normal control subjects, in the 1000 Genomes database, or in Exome Aggregation Consortium database, representing 60,662 individuals. Homozygous mutations in orthologous *GpnmB* have been previously reported to be associated with pigment dispersion syndrome (PDS) in DBA/2J mice. However, mutations in *GPNMB* have not yet been reported in humans with PDS. Our results indicate that mutations in *GPNMB* lead to ACD in humans.