Molecular outcomes, clinical consequences, and genetic diagnosis of Oculocutaneous Albinism in Pakistani population

Mohsin Shahzad1,*, Sairah Yousaf2,*, Yar M. Waryah3, Hadia Gul4, Tasleem Kausar2, Nabeela Tariq2, Umair Mahmood1, Muhammad Ali5, Muzammil A. Khan4, Ali M. Waryah3, Rehan S. Shaikh2, Saima Riazuddin1, Zubair M. Ahmed1 & University of Washington Center for Mendelian Genomics†

Nonsyndromic oculocutaneous Albinism (nsOCA) is clinically characterized by the loss of pigmentation in the skin, hair, and iris. OCA is amongst the most common causes of vision impairment in children. To date, pathogenic variants in six genes have been identified in individuals with nsOCA. Here, we determined the identities, frequencies, and clinical consequences of OCA alleles in 94 previously unreported Pakistani families. Combination of Sanger and Exome sequencing revealed 38 alleles, including 22 novel variants, segregating with nsOCA phenotype in 80 families. Variants of TYR and OCA2 genes were the most common cause of nsOCA, occurring in 43 and 30 families, respectively. Twenty-two novel variants include nine missense, four splice site, two non-sense, one insertion and six gross deletions. In vitro studies revealed retention of OCA proteins harboring novel missense alleles in the endoplasmic reticulum (ER) of transfected cells. Exon-trapping assays with constructs containing splice site alleles revealed errors in splicing. As eight alleles account for approximately 56% (95% CI: 46.52–65.24%) of nsOCA cases, primarily enrolled from Punjab province of Pakistan, hierarchical strategies for variant detection would be feasible and cost-efficient genetic tests for OCA in families with similar origin. Thus, we developed Tetra-primer ARMS assays for rapid, reliable, reproducible and economical screening of most of these common alleles.

Autosomal recessive non-syndromic oculocutaneous albinism (nsOCA) is a genetic disorder characterized by the partial or complete loss of pigmentation in the skin, hair, and iris. OCA is amongst the most common causes of vision impairment in children. To date, pathogenic variants in six genes, TYR, OCA2, TYRP1, SLC45A2, SLC24A5, and C10orf11, have been identified in individuals with nsOCA. We previously mapped a new locus for nsOCA on chromosome 4q24 in a large consanguineous Pakistani family, for which the gene is currently unknown. In published population studies, however, the detection rate of alleles causing albinism varies from 60% to 90%6–11.

The clinical range of OCA varies, with OCA1A being the most severe type, characterized by a complete lack of melanin production throughout life. The less severe forms (OCA1B, OCA2, OCA3, and OCA4) show some pigmentation over a lifetime. Although the diverse types of OCA are caused by variants in different genes, their clinical phenotypes are not always distinguishable, making molecular diagnosis an essential tool for genetic counseling12, and for emerging therapeutic interventions13. Our goals here were to determine identities, frequencies,
and clinical consequences of OCA alleles in Pakistani cohort, predominantly enrolled from Punjab province, and to develop hierarchical strategies for rapid, feasible, and cost-efficient genetic diagnostic assays for improved carrier detection and genetic counseling.

Results

After institutional review board approval, we enrolled 94 Pakistani families segregating nsOCA (Supplementary Fig. S1). All the affected individuals of these families have congenital hypopigmentation phenotype. Inter-familial variation of hair color was noted among individuals, ranging from white to honey blonde or brown (Table 1). Similarly, variation in iris color was noted, with tones ranging from light grey to brown (Table 1). However, due to limited facilities available in the remote areas of Pakistan, detailed clinical evaluations in every affected person from every family was not possible, therefore, we refrained ourselves from commenting on genotype-phenotype correlation for every variant. Through the combination of Sanger and whole exome sequencing techniques, we identified 38 variants, including 22 novel variants, segregating with the phenotype in 80 families (Table 2). The 22 new variants include 9 missense, 4 splice site, 2 non-sense, 1 insertion, and 6 gross deletions (Table 2 and Fig. S2). None of these variants were found in ethnically matched control samples (Table 2 and Table S1). We also documented the frequencies of polymorphic alleles of nsOCA genes in our cohort (Table S2).

Missense variants altered the targeting of encoded OCA proteins. The nine novel missense variants include two alleles of TYR (OCA1), six of OCA2 (OCA2) and one in TYRP1 (OCA3). Three prediction programs, specifically, Polyphen-2, MutationTaster and SIFT, suggested that each of these missense variants were deleterious (Table 2). To further evaluate the effects of these variants on the encoded proteins, we performed in silico molecular modeling and in vitro protein targeting studies.

TYR encodes tyrosinase in melanocytes, an essential enzyme for the biosynthesis of melanin. Previously, it was shown that missense alleles of tyrosinase lead to ER retention of encoded protein due to misfolding. To evaluate the targeting of p.Cys55Ser and p.Asp75Tyr harboring proteins, we introduced these variants in a GFP-tagged tyrosinase and transiently transfected human melanocytes. Wild type tyrosinase was localized throughout the cytoplasm of melanocytes (Fig. 1). Immunofluorescence studies with calregulin (an ER marker) demonstrated that the mutant proteins, however, predominantly co-localized with calregulin, indicating retention in the ER (Fig. 1). In contrast, p.(Asp86Tyr) missense variant did not affect the targeting of TYRP1 protein (Fig. 1).

OCA2 protein, with 12 putative transmembrane α-helices, belongs to the Na+/H+ antiporter family. Besides its presumed role in maintaining the pH of the melanosomes, OCA2 also participates in the sorting and transport of tyrosinase and tyrosinase-related protein 1 (TYRP1) to the plasma membrane. We performed comparative computational modeling of wild type and six novel missense alleles harboring OCA2 proteins using Phyre2 program. All the identified missense alleles were predicted to impact either protein folding, interaction with the lipid-bilayer, protein topology, or protein-protein interactions (Fig. S3). When transiently transfected in HEK293 cells, in contrast to wild type, disease-associated alleles of OCA2 proteins showed retention in the ER (Fig. 2). Collectively, these studies support the deleterious nature of novel missense variants identified in our OCA cohort.

Ex vivo splicing is defective due to splice site variants identified in Pakistani families. To evaluate the impact of four new splice site variants elucidated in our cohort, we examined the RNA splicing pattern of wild type and mutated exons by transfecting minigene constructs in COS7 cells. Results of our splicing assays are summarized in Fig. 3. The c.1037–18 T > G variant in exon 3 of TYR generated an upstream cryptic splice acceptor site, which resulted in retention of 55 bp from the coding region (Fig. 3). Retention of 17 base intronic region in the spliced product would result in the frameshift and premature truncation of the encoded tyrosinase. In contrast, the c.1184 + 2 T > G change in exon 3 of TYR revealed loss of canonical splice donor site. Splicing assay revealed utilization of a cryptic donor site within exon 3, resulting in loss of 35 bp from the coding region (Fig. 3), predicted to cause frameshift and premature truncation of the encoded protein. Both splice site variants of OCA2 (c.1182 + 2 T > TT and c.1951 + 4 A > G) revealed skipping of their respective exons in the minigene splicing assays (Fig. 3). The skipping of exon 11 (66 bp) due to c.1182 + 2 T > TT leads to in-frame deletion of 22 amino acids, while the loss of exon 18 (109 bp) of OCA2 due to c.1951 + 4 A > G variant would cause frameshift and premature truncation.

Exome sequencing revealed six gross deletions in OCA families. Approaches to detect indels using exome sequence data are an active area of research. As yet there is no single method that guarantees consistent success. We used the widely-evaluated methods XHMM and CoNIFER to identify gross insertions/deletions in our exome data. Our analyses revealed six novel gross deletions of TYR and OCA2 genes segregating with OCA phenotype in six families (Fig. 4). To investigate the mechanisms involved in these deletions, the intervals surrounding the breakpoints were analyzed through RepeatMasker. In addition, significantly over-represented motifs within ±15 bp of GRaBD translocations breakpoints were also sought. Thorough bioinformatics analysis revealed that the breakpoint for these deletions lie in the repetitive element, and the highly similar Alu short interspersed nuclear elements (SINEs) may serve as the substrate for nonhomologous recombination.

Next, to analyze the impact of deletions on protein 3D structure, we used Phyre2 modeling software. Removal of p.Ser395 to p.Leu529 amino acids due to deletion of exons 4 and 5 would eliminate the tyrosinase central domain that binds to its copper ligand for subsequent function (Fig. S4A). Similarly, deletions of OCA2 exons...
| Families | ID  | Sex | Allele          | Ethnicity      | Skin color | Hair color | Iris color | Visual Acuity | Right Refraction error | Fundus | Foveal Hypoplasia | Photophobia | Nystagmus |
|----------|-----|-----|-----------------|----------------|------------|------------|------------|---------------|------------------------|--------|-------------------|-------------|-----------|
| PKAB115  | IV:4| M   | c.164G>C        | Jatt sidhu     | White      | Yellow white | Grey blue | NA            | NA                     | NA     | NA                | NA          | Negative  |
| PKAB174  | IV:4| M   | c.223G>T        | Qureshi        | White      | Yellow white | Grey blue | NA            | NA                     | NA     | NA                | NA          | Positive  |
| PKAB200  | IV:4| M   | c.585G>A        | Gondal         | White      | White       | Grey blue | NA            | NA                     | NA     | NA                | NA          | Negative  |
| PKAB66   | V:3 | M   | c.649C>T        | Ali            | White      | Yellow white | Brown     | NA            | NA                     | NA     | NA                | NA          | Positive  |
| PKAB76   | IV:7| M   | c.649C>T        | Mochi          | White      | Yellow white | Brown     | NA            | NA                     | NA     | NA                | NA          | Positive  |
| PKAB141  | IV:3| M   | c.649C>T        | Arain          | White      | Yellow white | Hazel green | NA            | NA                     | NA     | NA                | NA          | Positive  |
| PKAB142  | IV:3| M   | c.649C>T        | Minhas         | White      | Yellow white | Hazel green | NA            | NA                     | NA     | NA                | NA          | Positive  |
| PKAB147  | III:3| M   | c.649C>T        | Raajpoot       | White      | Yellow white | Hazel green | NA            | NA                     | NA     | NA                | NA          | Positive  |
| PKAB118  | IV:2| F   | c.832C>T        | Mula khel      | White      | White       | Grey blue | 6/12          | 6/12                   | 1.00/-2.5*160/-1.00*160 | Albinotic | Positive | Positive | Positive |
| PKAB120  | III:13| M  | c.832C>T       | Arain          | White      | Grey         | 6/60      | 6/60         | 4.00*180–3.50*180       | Albinotic | Positive | Positive | Positive |
| PKAB143  | IV:7| M   | c.832C>T        | Malik          | White      | Grey         | 6/60      | 6/60         | Hyperopic              | Albinotic | Positive | Positive | Positive |
| PKAB145  | IV:2| F   | c.832C>T        | Arain          | White      | Yellow white | Grey blue | NA            | NA                     | NA     | NA                | NA          | NA        |
| PKAB165  | IV:8| M   | c.832C>T        | Jatt           | White      | Grey blue    | 6/60      | 6/60         | Hyperopic              | Albinotic | Positive | Positive | Negative |
| PKAB183  | VI:1| F   | c.1037–7T>A     | Sayyad bukhari | Pinkish    | Brownish    | 6/60      | 6/60         | Hyperopic              | Albinotic | Positive | Positive | Positive |
| PKAB185  | IV:4| F   | c.832C>T        | Sayyad bukhari | White      | White       | Grey blue | 6/60          | 6/60                   | Hyperopic | Albinotic | Positive | Positive |
| PKAB186  | IV:3| F   | c.1037–7T>A     | Sayyad bukhari | Yellow     | Pinkish     | Brownish  | 6/60          | 6/60                   | Hyperopic | Albinotic | Positive | Positive |
| PKAB188  | VI:1| M   | c.1037–7T>A     | Khokar         | Yellow     | Pinkish     | Brownish  | 6/60          | 6/60                   | Hyperopic | Albinotic | Positive | Positive |
| PKAB189  | VIII:11| M   | c.832C>T       | Bhatti         | White      | White       | Grey blue | NA            | NA                     | NA     | NA                | NA          | NA        |
| PKAB194  | VII:2| M   | c.832C>T       | Raajpoot       | White      | White       | Grey blue | NA            | NA                     | NA     | NA                | NA          | NA        |
| LUAB05   | IV:8| F   | c.832C>T       | Urdu speaking  | Pinkish     | White       | Grey blue | 6/60          | 6/60                   | Myopic type | Pigmented | Poorly formed | Positive | Positive |
| LUAB08   | V:7 | M   | c.832C>T       | Urdu speaking  | White       | White       | Grey blue | 4/60          | 4/60                   | Clear     | Absent            | Positive | Negative |
| LUAB19   | IV:3| M   | c.832C>T       | Urdu speaking  | White       | White       | Grey blue | NA            | NA                     | NA     | NA                | NA          | NA        |
| PKAB204  | IV:9| M   | c.832C>T       | Jatt           | White      | Grey blue   | 6/60      | 6/60         | NA                     | NA     | NA                | NA          | Positive |
| PKAB128  | IV:3| M   | c.896G>A       | Abbasi         | White      | Grey blue   | 6/60      | 6/60         | NA                     | NA     | NA                | NA          | NA        |
| PKAB154  | IV:5| M   | c.896G>A       | Warase         | White      | Grey blue   | 6/60      | 6/60         | NA                     | NA     | NA                | NA          | NA        |
| PKAB138  | III:7| F   | c.943–948delTCAGCT | Butt       | White      | White       | Grey blue | 6/12          | 6/18                   | 3.00*160–0.50*160 | Albinotic | Positive | Positive | Positive |
| LUAB01   | IV:4| M   | c.1037–18T>G    | Urdu speaking  | White       | Light brown | 6/60      | 6/60         | Hypermetropia           | Clear    | Poorly formed | Positive | Positive |
| PKAB105  | IV:9| M   | c.1037–7T>A     | Gouraya        | White      | Yellow white | NA        | NA            | NA                     | NA     | NA                | NA          | Negative |
| PKAB197  | IV:3| M   | c.1037G>A       | Sahoo          | White      | Grey blue   | NA        | NA            | NA                     | NA     | NA                | NA          | NA        |
| PKAB119  | IV:9| M   | c.1147G>A       | Bangulzai      | White      | Grey/ brown | 6/12      | 6/18         | 1.50/-3.00*100/-1.50*300*180 | Albinotic | Positive | Positive | Positive |
| PKAB130  | IV:8| F   | c.1184+2T>C     | Bhatti         | White      | Light green | 6/60      | 6/60         | Myopic astigmatism      | Albinotic | Positive | Positive | Positive |
| PKAB193  | IV:2| F   | c.1204C>T      | Gujar          | White      | White       | Grey blue | NA            | NA                     | NA     | NA                | NA          | NA        |
| PKAB64   | III:6| M   | Exons 4–5 deletion | Jatt         | Golden Blonde | Brownish green | 6/60   | 6/60        | +1.75*160–2.75*2180    | —       | Present           | Present    | Present  |
| PKAB168  | IV:8| M   | Exons 4–5 deletion | Chakotaray   | Grey white  | White       | Brown     | NA            | NA                     | NA     | NA                | NA          | Positive |
| PKAB81   | IV:5| F   | c.1255G>A      | Raiport        | White      | Yellow white | Grey      | 6/60          | 6/24                   | Hyperopic | Albinotic | Positive | Positive |

Continued
| Families | ID | Sex | Allele | Ethnicity | Skin color | Hair color | Iris color | Visual Acuity | Type of refraction error | Fundus | Forced hypoplasia | Photophobia | Nystagmus |
|----------|----|-----|--------|-----------|------------|------------|------------|----------------|------------------------|--------|----------------|------------|-----------|
| PKAB108  | IV:4 | M  | c.1255 G > A | Shaikh | White | Golden yellow | Grey blue | NA | NA | NA | NA | NA | NA |
| PKAB107  | IV:1 | F  | c.1255 G > A | Chanarr | White | Yellow white | Grey blue | 6/60 | 6/60 | Hyperopic | Albinotic | Positive | Positive | Positive |
| PKAB140  | IV:7 | F  | c.1255 G > A | Daaye | White | Yellow white | Grey blue | 6/18 | 6/18 | Myopic | Albinotic | Positive | Positive | Positive |
| PKAB184  | IV:6 | M  | c.1255 G > A | Oteera | White | Grey blue | CF | NA | Albinotic | Positive | Positive | Positive | |
| PKAB191  | V:4  | M  | c.1255 G > A | Dodiyanar | White | Grey blue | NA | NA | NA | NA | NA | NA | NA |
| LUAB03   | V:6  | M  | c.1255 G > A | Sindi | Pinkish white | White | Hazel | 1/60 | 1/60 | No improvement | Clear | Poorly formed | Positive | Positive |
| GUAB01   | IV:6 | M  | c.1255 G > A | Saraiki speaking | Light blonde | Light blonde | Blue grey | NA | NA | NA | NA | NA | Positive | Positive |

**OCA2**

| Families | ID | Sex | Allele | Ethnicity | Skin color | Hair color | Iris color | Visual Acuity | Type of refraction error | Fundus | Forced hypoplasia | Photophobia | Nystagmus |
|----------|----|-----|--------|-----------|------------|------------|------------|----------------|------------------------|--------|----------------|------------|-----------|
| PKAB70   | IV:7 | M  | Exons 3–14 deletion | Maachi | White | Yellow | Green | 6/36 | 6/36 | Hyperopic | Albinotic | Positive | Positive | Positive |
| PKAB69   | V:6  | M  | Exons 7–8 deletion | Mughal | Pink white | Silver yellow | Grey | CF | CF | −3.00/−3.50*178–3.00/−4.00*3 | Albinotic | Positive | Positive | Positive |
| PKAB77   | IV:6 | M  | Exons 7–8 deletion | Terkhan | Pink white | Silver yellow | Grey | 6/36 | 6/36 | −1.75/−1.00*5.25/−1.25*20 | Albinotic | Positive | Positive | Positive |
| PKAB127  | IV:4 | M  | c.1045–15T > G | Rajpoott | White | Yellow | Hazel brown | 6/48 | 6/48 | Compound astigmatism | Albinotic | Positive | Positive | Positive |
| PKAB132  | III:5 | M  | c.1045–15T > G | Jatt | White | Yellow | Light green | NA | NA | NA | NA | NA | NA | |
| PKAB134  | III:3 | M  | c.1045–15T > G | Abbasi | White | Yellow | Light green | NA | NA | NA | NA | NA | NA | |
| PKAB135  | IV:3  | M  | c.1045–15T > G | Arain | White | White | Light green | NA | NA | NA | NA | NA | NA | |
| PKAB177  | IV:7  | F  | c.1045–15T > G | Bhatti | White | Yellow | Hazel brown | 6/60 | 6/60 | Hypermetropia | Pigmented | Positively | Pigmented | Positive |
| GUAB03   | IV:8  | F  | c.1045–15T > G | Saraiki speaking | Brown | Brown | Greenish | NA | NA | NA | NA | Positive | Positive | Positive |
| LUAB07   | IV:3  | M  | c.1045–15T > G | Sindi speaking | White | White | Light brown | 6/18 | 6/18 | Hypermetropia | Pigmented | Positively | Pigmented | Positive |
| LUAB02   | IV:8  | F  | c.1056 A > C | Sindi speaking | White | White | Light | 3/18 | 3/18 | Hypermetropia | Pigmented | Positively | Pigmented | Positive |
| PKAB136  | IV:3  | F  | c.1064 C > T | Chandia | White | Yellow white | Light green | NA | NA | NA | NA | NA | NA | |
| GUAB02   | IV:6  | M  | c.1075 G > C | Saraiki speaking | Light golden | Light golden | Greenish | NA | NA | NA | NA | Positive | Positive | |
| PKAB144  | IV:2  | M  | c.1182+2T > TT | Arain | White | Yellow | Grey blue | NA | NA | NA | NA | NA | NA | |
| PKAB148  | IV:5  | M  | c.1211 C > T | Sayyad | White | Golden yellow | Grey blue | NA | NA | NA | NA | NA | NA | |
| LUAB06   | IV:8  | M  | c.1322 A > G | Sindi speaking | Pinkish white | Golden white | Light brown | 6/60 | 6/60 | Myopic Type | Pigmented | Positively | Pigmented | Positive |
| PKAB172  | V:1  | F  | c.1456 G > T | Wains | White | Yellowish brown | Brown | 6/48 | 6/40 | Myopic astigmatism | Albinotic | Positive | Positive | Positive |
| PKAB187  | III:3 | M  | c.1456 G > T | Bhutta | White | Golden yellow | Grey blue | NA | NA | NA | NA | NA | NA | |
| GUAB06   | IV:5  | M  | c.1456 G > T | Saraiki speaking | Golden | Golden | Greenish | NA | NA | NA | NA | NA | Positive | Positive |
| LUAB17   | IV:3  | F  | c.1456 G > T | Sindi speaking | White | Yellow white | Grey blue | NA | NA | NA | NA | NA | NA | |
| PKAB122  | IV:2  | F  | c.1922 C > T | Mughal | White | Golden yellow | Light brown | 6/24 | 6/24 | NA | Albinotic | Positive | Positive | Positive |
| LUAB10   | III:9 | F  | c.1951+4A > G | Urdu speaking | White | Light brown | 6/18 | 6/18 | Hypermetropia | Pigmented | Positively | Pigmented | Positive |
| PKAB137  | IV:4  | M  | Exon 19 deletion | Malick | White | Yellow white | Hazel green | NA | NA | NA | NA | NA | NA | Positive |
| PKAB121  | IV:3  | F  | Exons 20–24 deletion | Achakzai | White | Brown | 6/18 | 6/18 | 2.50*20/2.0*160 | Albinotic | Positive | Positive | Negative |
| PKAB116  | IV:4  | M  | Exons 22–24 deletion | Raipoot | White | Golden yellow | Misty grey | 6/36 | 6/36 | +3.75/−4.75*15 | +1.75/−3.0*80 | Albinotic | Positive | Positive | Positive |

Continued
| Families | ID   | Sex | Allele   | Ethnicity        | Skin color | Hair color | Iris color | Visual Acuity (Right | Visual Acuity (Left) | Type of refraction error | Fundus | Foveal hypoplasia | Photophobia | Nystagmus |
|----------|------|-----|----------|------------------|------------|------------|------------|----------------------|------------------------|----------------------------|---------|------------------|-------------|-----------|
| LUAB09   | IV:5 | F   | c.2228 C > T | Urdu speaking | White       | Goldish white | Light brown | 6/18 | 6/18                | Hypermetropia            | Pigmented                        | Poorly formed | Positive | Positive |
| PKAB170  | V:3  | F   | c.2360 C > A | Jatt Langa      | White       | Yellow      | Brown      | 6/36 | 6/36                | Hyperopic                | Albinotic                        | Positive | Positive | Positive |
| PKAB125  | IV:5 | M   | c.2360 C > T | Arain           | White       | Golden yellow | NA         | NA | NA                  | NA                      | NA                                | NA       | NA         | Positive |

**TYR1**

| PKAB166  | IV:3 | F   | c.256 G > T | Bhutta         | Brown       | Wheat      | Brown | 6/6 | 6/6                | Emmetropic              | Normal                        | Negative | Negative | Positive |
| PKAB164  | IV:5 | M   | c.1067 G > A | Langa          | Grey brown  | Wheat      | Brown | 6/24 | 6/24                | −2.5/−5.25/−97/0.50/−5.00/−90 | Albinotic                     | Positive | Positive | Positive |
| PKAB192  | V:2  | M   | c.1067 G > A | Harajaj        | Wheat       | Golden yellow | Light green | NA | NA                  | NA                      | NA                                | NA       | NA         | Positive |
| PKAB205  | IV:10| F   | c.1067 G > A | Matam          | Wheat       | Golden yellow | Light green | NA | NA                  | NA                      | NA                                | NA       | NA         | Positive |
| PKAB146  | III:1| M   | c.1532 C > T | Shaikh         | Wheat       | Yellow brown | Grey     | 6/9 P | 6/9 P                | −8.0/−1.0/10/7.50/−1.00/180 | Normal                        | Negative | Positive | Positive |

**SLC45A2**

| GUAB04   | IV:5 | M   | c.1532 C > T | Saraiki speaking | Light brown | Golden | Blush | NA | NA                | NA                      | NA                                | NA       | Positive | Positive |
| GUAB05   | II:9 | M   | c.1532 C > T | Saraiki speaking | Light brown | Golden | Blush | NA | NA                | NA                      | NA                                | NA       | Positive | Positive |

Table 1. Clinical features of oldest affected individuals of families with variants in OCA1-4. NA: Not available.

would result in the partial or complete loss of Na-Sulphur-symporter domain, which mediates the intake of several different molecules with the concomitant uptake of Na⁺ (Fig. S4B) and thus predicted to result in non-functional, truncated proteins.

**Alleles of TYR and OCA2 are the common cause of nsOCA in our cohort.** Variants of TYR (allele frequency: 15/38) and OCA2 (19/38) were the most common cause of nsOCA, occurring in 43 and 30 families, respectively (Fig. 5A). To further refine the prevalence estimates, we reviewed the alleles frequencies in known albinism genes among our larger cohort of 143 families, including 48 previously reported families (Fig. 5B) and in published literature (Table S1). TYR and OCA2 alleles are the frequent cause of OCA in Pakistanis (Table S1). In our cohort, variants in TYR and OCA2 collectively account for the majority [67.8% (97/143); 95% confidence interval (CI): 60.2–76.0%] of the genetic causes of nsOCA, which is comparable to prevalence in European population (Table S3). In approximately 14% of our OCA families, we did not find any pathogenic variant in the known OCA genes (Fig. 5B).

Next, we investigated the frequency of alleles of nsOCA genes in our cohort. Overall, four alleles of TYR, three of OCA2, and one of SLC45A2 together account for ~56% (95% CI: 46.52–65.24%) of the variants responsible for nsOCA in our cohort (Fig. 5C). Therefore, we developed rapid and inexpensive assays for detecting carriers and homozygotes (Fig. 5D). For most of these alleles, we were able to develop tetra-primer ARMS assays. The sensitivity and specificity of these assays were confirmed on multiple DNA samples with different genotypes followed by Sanger sequencing.

**Discussion**

Our study illustrates the relative genetic contribution of four major OCA genes in the prevalence of albinism, primarily in families (69) currently residing in the Punjab province of Pakistan. However, our study also includes 19 families from Sindh and 6 families from Khyber Pakhtunkhwa (KPK) provinces. Pakistan has a rich anthropogenic background owing to successive waves of invasions and adaptations of haplogroups. Most did not intermingle with the original local population and practiced endogamy, giving rise to genetic isolates that persist even today. Parental consanguinity is an important risk factor (0.25–20% higher chances) for recessive genetic defects34. In Pakistan, 62.7% of marriages are consanguineous, ~80% of which are between first cousins35. Persistence of endogamy has resulted in the enrichment of a particular genetic order (Table 1). For instance, c.832 C > T in TYR and c.1045–15 T > G in OCA2 are the most frequent alleles observed in our samples (Fig. 5C). However, both these alleles were observed in families of various ethnic and geographical origins (Table 1). Similarly, other common variants (e.g. c.649 C > T, c.1255 G > A, c.1456 G > T) were also found in families enrolled from Punjab, Sindh and KPK (Table 1). Therefore, with the current samples size for each of the identified allele, it would be inapplicable to comment on ethnic/linguistic origins of different alleles observed in our cohort.
| Families     | Variant                      | Effect on protein | Allele frequency |
|--------------|------------------------------|-------------------|------------------|
| PKAB115      | c.164 G > C                 | p.(Cys55Ser)      | PK 0/180 0.000   |
| PKAB174      | c.223 G > T                 | p.(Asp75Tyr)      | 0 0 0 8.2        |
| PKAB200      | c.585 G > A                 | p.(Trp195*)       | 0/186 0 1.53 8.2 |
| PKAB66, PKAB76, PKAB141, PKAB142, PKAB147 | c.649 C > T | p.(Arg217Tyr) | — 0 1.53 197.9 |
| PKAB118, PKAB120, PKAB143, PKAB145, PKAB165, PKAB185, PKAB189, PKAB194, LUAB05, LUAB08, LUAB19, PKAB204 | c.832 C > T | p.(Arg278*) | — 0 0 190.0 |
| PKAB128, PKAB154 | c.896 G > A   | p.(Arg299His)    | — 0 0 74.2       |
| PKAB138      | c.943-946delTCAGCT         | p.(313-316delSerAla) | — 0 0 0 |
| PKAB105, PKAB183, PKAB186, PKAB188 | c.1037–7 T > A | Splicing error | — 0 15.38 0 |
| LUAB01       | c.1037–18 T > G            | Splicing error    | 0/176 0 0 16.7   |
| PKAB197      | c.1037 G > A               | p.(Gly346Glu)     | 0 0 16.7         |
| PKAB119      | c.1147 G > A               | p.(Asp383Asn)     | — 0 0 165        |
| PKAB130      | c.1184 + 2 T > C           | Splicing error    | 0/178 0 0 8.2    |
| PKAB193      | c.1204 C > T               | p.(Arg402*)       | — 0 0 49.84      |
| PKAB64, PKAB168 | Exons 4–5 deletion  | Frame shift | 0/180 0 0 |
| PKAB81, PKAB106, PKAB107, PKAB140, PKAB184, PKAB191, LUAB03, LUAB08, LUAB19, GUAB01 | c.1255 G > A | p.(Gly419Arg) | 0 0 33.12 |
| PKAB070      | Exons 3–14 deletion        | Frame shift       | 0/176 0 0 0      |
| PKAB069, PKAB77 | Exons 7–8 deletion        | Frame shift       | 0/180 0 0 0      |
| LUAB04       | c.987 C > AGA              | p.(Gln330Aspfs*2) | 0/184 0 0 0      |
| PKAB127, PKAB132, PKAB133, PKAB134, PKAB135, PKAB177, LUAB07, GUAB03 | c.1045–15 T > G | Splicing error | 0 41.36 |
| LUAB02       | c.1056 A > C               | p.(Arg352Ser)     | 0/182 0 0 0      |
| PKAB136      | c.1064 C > T               | p.(Ala355Val)     | 0/184 0.001 289.2 |
| GUAB02       | c.1075 G > C               | p.(Gly359Arg)     | 0/176 0 0 0      |
| PKAB144      | c.1182 + 2 T > TT          | Splicing error    | 0/180 0 0 0      |
| PKAB148      | c.1211 C > T               | p.(Thr404Met)     | — 0 7.69 74.17   |
| LUAB06       | c.1322 A > G               | p.(Asp441Gly)     | 0/178 0 1.53 16.7 |
| PKAB172, PKAB187, LUAB06, LUAB17 | c.1456 G > T | p.(Asp486Tyr) | — 0 0 16.7 |
| PKAB122      | c.1922 C > T               | p.(Ser641Leu)     | 0/184 0 0 0      |
| LUAB10       | c.1951 + 4 A > G           | Splicing error    | 0 0 0 0          |
| PKAB137      | Exon 19 deletion           | Frame shift       | 0/174 0 0 0      |
| PKAB121      | Exons 20–24 deletion       | Frame shift       | 0/176 0 0 0      |
| PKAB116      | Exons 22–24 deletion       | Frame shift       | 0/180 0 0 0      |
| LUAB09       | c.2228 C > T               | p.(Pro743Leu)     | — 0 0 90.78      |

Continued
Our overall estimated prevalence of TYR and OCA2 alleles is quite similar in certain cases from Europe but fairly different from other studies in different populations (Table S3). For instance, TYR and OCA2 variants account for 70% and 10%, respectively, of OCA in a study of 127 patients from a Chinese population (Table S3). In a few eastern and central regions of China, TYR and OCA2 contribution varies; however, TYR alleles remain the most common cause of OCA. In India, a study of 82 OCA patients revealed approximately 60% and 11% prevalence for TYR and OCA2, respectively36. Similarly, in the US, Europe, Italy, Japan, and Korea, the alleles of TYR are the most common cause of nsOCA7–9,11,37 (Table S3). In contrast, variants in TYR account for ~80% of the OCA cases in an African population38.

Usually the alleles of TYR result in the retention of encoded tyrosinase in the ER18,30. Here, we also observed retention of OCA2 proteins harboring missense and truncating alleles in the ER (Fig. 2). A portion of the known human and mouse TYR variants have shown temperature-sensitive behavior30,39–44. Cultivating mammalian cells at permissive temperature (31 °C) resulted in increased cytoplasmic expression of tyrosinase harboring missense alleles, especially for those present in the copper-binding region30,39,41,42,44. However, the new OCA2 alleles were imaged at 31 °C and did not reveal an apparent increased cytoplasmic expression (data not shown).

Besides single nucleotide variants, our study, for the first time in Pakistani population, revealed six novel gross deletions in TYR and OCA2 genes (Fig. 4). These genetic aberrations span from a single exon deletion up to 12 exons with the encompassed introns (Fig. 4). These genetic aberrations confer pigmentation disorders either because of almost entire protein would be missing or non-functional truncated forms would be encoded. Considering their relative contribution simplified methods and bioinformatics tools will be needed for rapid detection of gross deletions for clinical genetic diagnosis of OCA.

Clinical presentation of OCA is often not very helpful in genetic diagnosis due to significantly overlapping features (Table 1). For economic and geographical reasons, it is not feasible to routinely perform Sanger sequencing of all the known OCA genes to detect underlying genetic defects. However, PCR-based assays that are developed to detect common alleles are quick and affordable. Twenty of the variants identified in this study were found in single families. However, eight variants (Fig. 5C) account for more than half of the alleles found in our families (Table 2). Therefore, we developed tetra-primer ARMS assays for these common alleles for rapid detection and genetic screening in large cohorts before embarking for exome/genome wide studies. We are cognizant of the fact that the majority of families in our cohort are from the Punjab province of Pakistan. Hence, the details of a hierarchical nonsyndromic OCA genes mutational screening strategy may need to be refined for other geographical regions and lingo-ethnic groups within Pakistan. Our results will be helpful for future diagnosis, genetic counseling, molecular epidemiology, and functional studies of nsOCA genes associated pigmentation disorders.

Intriguingly, fourteen families in our cohort did not reveal any pathogenic variants in common nsOCA genes. There are several possibilities for our failure to detect disease-associated alleles in these fourteen families. First, potential pathogenic variants may alter sequence of cis-acting regulatory or deep intronic splicing elements that are necessary for expression of these genes in melanocytes. Presently, we have very limited knowledge of the location of the regulatory elements of nsOCA genes. Secondly, the hypopigmentation segregating in these fourteen families may be syndromic and requires comprehensive clinical phenotyping, which would help in filtering the candidate genetic variants for further assessments. Third, there may be additional genes responsible for OCA in humans. For instance, in the US population the detection rate of alleles in known nsOCA genes is around 75%10. Thus, there exists both a need for further genetic understanding of albinism and an opportunity to improve the molecular diagnosis of albinism, and quite possibly its prevention.

**Methods**

**Patients.** This study followed the tenets of the Declaration of Helsinki. This study was approved by the IRB Committees at the University of Maryland School of Medicine (UMSOM), USA, the Institute of Molecular Biology & Biotechnology, Multan, Liaquat University of Medical & Health Sciences, Jamshoro, and Gomal Centre

| Families | Variant | Effect on protein | Allele frequency |
|----------|---------|------------------|-----------------|
|          |         |                  | PK 1000 genome | NHLBI10<sup>+</sup> | ExAC 10<sup>+</sup> | Polyphen 2 | Mutation Taster | SIFT |
| PKAB170  | c.2360 C > A  | p.(Ala787Glu)   | 0/182         | 0          | 0          | 0     | Probably damaging | Disease causing | Damaging |
| PKAB125  | c.2360 C > T  | p.(Ala787Val)   | —             | 0          | 0          | 24.78 | Probably damaging | Disease causing | Damaging |
| TYRP1    | PKAB166  | c.256 G > T     | p.(Asp86Tyr)   | 0/180       | 0          | 0     | Probably damaging | Disease causing | Damaging |
|          | PKAB164, PKAB192, PKAB205 | c.1067 G > A | p.(Arg356Gln) | —           | 0          | 0     | 25.50 | Probably damaging | Disease causing | Damaging |
| PKAB146  | c.1534 C > T  | p.(Glu512<sup>*</sup>) | 0/184       | 0          | 0          | 16.53 | —           | —           | —           |
| SLC45A2  | GUAB04, GUAB05 | c.1532 C > T  | p.(Ala511Val) | —           | 0          | 0     | 8.2     | Probably damaging | Disease causing | Damaging |

Table 2. Overview of molecular outcome and predictive effects of alleles in TYR, OCA2, TYRP1 and SLC45A2 genes. Novel mutations found in this study are in bold. (PK: Pakistani Control).
Figure 1. Subcellular distribution of wild type and mutant tyrosinase and tyrosinase-like proteins in human melanocytes. eGFP-tagged TYR and TYRP1 wild type and mutant constructs (green) were transiently transfected in melanocytes grown at 37 °C. Calregulin (red) and DAPI (blue) were used as markers for the endoplasmic reticulum and nucleus, respectively. Merged images show the co-localization of only tyrosinase variants with calregulin, which indicated ER retention. The scale bar represents 20 μm for all panels.
of Biochemistry and Biotechnology, Gomal University, D.I. Khan, Pakistan. All the methods were performed in accordance with the UMSOM relevant guidelines and regulations. Pedigrees were drawn after interviewing multiple individuals to confirm the relationships. Informed written consents were obtained from the adult subjects and the parents of minor subjects. Two to five ml of peripheral blood samples were collected from each participating individual. Human genomic DNA was isolated from peripheral blood by using inorganic method 45.

Detailed medical histories were taken for all of the participating individuals of the enrolled families, including the hypopigmentation phenotype of the hair, skin, eye, disease onset, segregation, presence of eye abnormalities (nystagmus, strabismus, photophobia, and poor vision) and information about immunological, neurological or bleeding time.

**Sanger sequencing and segregation analysis.** Primers (Table S4) covering the coding regions of the TYR, OCA2, TYRP1 and SLC45A2 genes were designed in primer3 software (http://bioinfo.ut.ee/primer3–0.4.0/primer3/input.htm) for Sanger sequencing and segregation analysis. PCR amplification, cleaning, and sequencing reactions were performed as previously described30,32.

**Exome Sequencing.** DNA samples from some of the participating family members were submitted for exome sequencing (WES) at the Center for Mendelian Genomics (CMG), University of Washington. WES on PKAB137 was performed at the Genetic core facility of Cincinnati Children’s Hospital. Genomic libraries were recovered for exome enrichment using the NimbleGen Exome kits (Roche Diagnostics; San Francisco, CA). Paired-end sequencing was performed using the Illumina Hi-Seq 2000 system (Illumina, San Diego, CA). The obtained sequencing data were analyzed following the guidelines outlined in the Broad Institute's Genome Analysis Toolkit46. The raw data were mapped using the Burrows Wheeler Aligner47. Variants were called using Unified Genotyper software, and the data were then subjected to further processing and quality control46,47. Golden Helix software was used to analyze the variants found through exome sequencing. The coverage of exome was checked through Golden Helix Genome Browser.

**Bioinformatics Analysis.** Pathogenicity of novel missense variants was assessed by different bioinformatics prediction programs; Polyphen-2, Sift, and Mutation Taster. Allele frequencies of identified variants were checked

---

**Figure 2.** Subcellular distribution of wild type and mutant OCA2 in HEK293 cells. eGFP-tagged OCA2 wild type and mutant constructs (green) were transiently transfected in HEK293 cells grown at 37 °C. Calregulin (red) and DAPI (blue) were used as markers for the endoplasmic reticulum and nucleus, respectively. Merged images show partial co-localization of OCA2 variants with calregulin, which indicated increased ER retention. The scale bar represents 20μm for all panels.
in 1000 Genome browser (http://browser.1000genomes.org/), NHLBI-ESP variant database (http://evs.gs.washington.edu/EVS/) and ExAC database (http://exac.broadinstitute.org/). Multiple sequence alignments of orthologous OCA proteins were performed by using Clustal Omega multiple alignment program (http://www.ebi.ac.uk/Tools/msa/clustalo/). The 3D protein structures were predicted by Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) with "Intensive Mode" option which is a combination of template-based modeling and ab initio methods.

**Tetra Primer Amplification Refractory Mutation System (ARMS) assay.** PCR primers (Table S5) used for Tetra-primers ARMS assay were designed by adjusting maximum (1:8) and minimum (1:3) relative size difference of two inner products and keeping others by default settings. Reaction mixture contains 10 ul 2X

---

**Figure 3. Functional analysis of splice site variants.** Schematic representation and Sanger sequencing chromatograms of minigene assays for TYR and OCA2 novel splice site variants revealed aberrant splicing products, supporting their pathogenic nature.
ECONOQAQ, 0.5 ul of 10 Um outer primers and 1.0 ul of 10 Um inner primers each, 50 ng (2 ul) genomic DNA and 5 ul H2O. The thermo cycling conditions are: initial denaturation of 4 min at 95 °C and followed by 35 cycles of 95 °C for 30 sec, 60 °C for 45 sec, 72 °C for 45 sec, and a final extension at 72 °C for 8 min. The PCR product was visualized in 2.0% agarose gel stained with ethidium bromide.

**Exon-trapping assay.** To ascertain the consequences of novel splice site variants found in our OCA families, the wild and mutant exons along with flanking intronic region (200 bp) were PCR amplified, cloned in pSPL3 vector (Invitrogen, Carlsbad, CA) and sequence verified as described. Purified cloned constructs were transfected into COS7 cells using PEI (Polyethyleneimine). After 48 hours of transfection, RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA) and single stranded cDNA (Clontech, Mountain View, CA) was synthesized. Primary PCR amplification of cDNA was performed using SD6 and SA2 vector primers and amplified products were cloned in TA-cloning vector (Invitrogen). At least ten bacterial clones for each construct were Sanger sequenced.

**Expression constructs, transfection and immunofluorescence.** All expression constructs used for immunofluorescence study were in eGFP-tagged vector (Clontech). For generating mutant constructs, mutagenesis was performed by using QuickChange kit (Stratagene, La Jolla, CA) and used specific wild type eGFP construct as a template. Each construct was transiently transfected in melanocytes or HEK293
cells seeded on cover slips with Lipofectamine 3000 (Invitrogen). After 48 hours of transfection either at 37 °C or 31 °C, cells were fixed and permeabilized in 4% paraformaldehyde and 0.1% Triton X-100 in PBS, respectively. For endoplasmic reticulum and nucleus visualization anti-Calregulin antibody (Santa Cruz Biotechnology, Santa Cruz, CA) and DAPI staining was performed. A Zeiss LSMDUO confocal microscope was used for imaging.

References
1. Spritz, R. A., Chiang, P. W., Oiso, N. & Alkhateeb, A. Human and mouse disorders of pigmentation. *Curr Opin Genet Dev* **13**, 284–289 (2003).
2. Tomita, Y. & Suzuki, T. Genetics of pigmentary disorders. *Am J Med Genet C Semin Med Genet* **131C**, 75–81, doi: 10.1002/ajmg.c.30036 (2004).
3. Oetting, W. S. & King, R. A. Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism. *Hum Mutat* **13**, 99–115, doi: 10.1002/humu.1026 (1999).
4. Montoliu, L. et al. Increasing the complexity: new genes and new types of albinism. *Pigment Cell Melanoma Res* **27**, 11–18, doi: 10.1111/j.1109-2739.2014.01682.x (2014).

Figure 5. Prevalence of albinism genes and their alleles in a Pakistani population. (A–C) Relative distribution of variants in albinism genes in Pakistani families. (A) The distribution of OCA1–4 genes in 94 families screened in the current study. Number of families and their percentage contribution are given in parentheses. (B) Total pool of albinism genes in our cohort. (C) Frequencies of recurrent alleles of mOCA genes in Pakistanis. (D) Results of tetra-primer ARMS assays for detection of recurrent variants. The top band in each gel represents the positive control amplimer in all samples generated using the outer primers. Nested allelespecific primers were used to generate the wild type (Wild) or variant harboring (Mutant) PCR products. IC: inner control product.
45. Grimberg, J. et al. A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. *Nucleic Acids Res* **17**, 8390 (1989).

46. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303, doi: 10.1101/gr.107524.110 (2010).

47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, doi: 10.1093/bioinformatics/btp324 (2009).

48. Kelley, I. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363–371, doi: 10.1038/nprot.2009.2 (2009).

49. Ahmed, Z. M. et al. Nonsyndromic recessive deafness DFNB18 and Usher syndrome type IC are allelic mutations of USH1C. *Hum Genet* **110**, 527–531, doi: 10.1007/s00439-002-0732-4 (2002).

**Acknowledgements**

We thank the participants for their cooperation. We thank Shakeel Sheikh, Fareeha Zulfiqar, Suruchi Shrestha and Thomas Jaworek for their technical assistance. We thank Drs. Saumil Sethna, Elodie Richard and Arnaud Giese for their valuable critiques of the work and helpful discussions. The authors acknowledge support of the National Human Genome Research Institute and the National Heart, Lung, and Blood Institute. Funding was provided by HG006504 (Yale Center for Mendelian Disorders), HG006542 (Baylor-Hopkins Center for Mendelian Genomics) and HG006493 (University of Washington Center for Mendelian Genomics). This study was supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases research grant R56AR065483 to ZMA.

**Author Contributions**

Conception and design of the work: M.A.K., A.M.W., R.S.S., S.R., Z.M.A. Acquisition of data: M.S., S.Y., T.K., U.W.C.M.G., Y.M.W., H.G., N.T., U.M. Analysis and interpretation of data: all authors. Drafting of the manuscript: M.S., S.Y., S.R., Z.M.A. Critical revision of the manuscript: S.R., Z.M.A. Final approval of the version to be published: all authors.

**Additional Information**

Supplementary information accompanies this paper at http://www.nature.com/srep

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Shahzad, M. et al. Molecular outcomes, clinical consequences, and genetic diagnosis of Oculocutaneous Albinism in Pakistani population. *Sci. Rep.* **7**, 44185; doi: 10.1038/srep44185 (2017).

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2017
Consortia
University of Washington Center for Mendelian Genomics (UW CMG) Consortium

Michael J. Bamshad⁶,⁷, Jay Shendure⁶, Deborah A. Nickerson⁶, Gonçalo R. Abecasis⁸, Peter Anderson⁶, Elizabeth Marchani Blue⁶, Marcus Annable⁶, Brian L. Browning⁶, Kati J. Buckingham⁶, Christina Chen⁶, Jennifer Chin⁶, Jessica X. Chong⁶, Gregory M. Cooper⁸, Colleen P. Davis⁸, Christopher Frazier⁶, Tanya M. Harrell⁶, Zongxiao He¹⁰, Preti Jain⁹, Gail P. Jarvik⁶, Guillaume Jimenez⁶, Eric Johanson⁶, Goo Jun⁸, Martin Kircher⁶, Tom Kolar⁶, Stephanie A. Krauter⁶, Niklas Krumm⁶, Suzanne M. Leal¹⁰, Daniel Luksic⁶, Colby T. Marvin⁶, Sean McGee⁶, Karynne Patterson⁶, Marcos Perez⁶, Sam W. Phillips⁶, Jessica Pijoan⁶, Christa Poel⁶, Seamus Ragan⁶, Frederic Reinier⁶, Peggy D. Robertson⁶, Regie Santos-Cortez¹⁰, Aditi Shankar⁶, Krystal Slattery⁶, Cindy Shephard⁶, Kathryn M. Shively⁶, Deborah L. Siegel⁶, Joshua D. Smith⁶, Holly K. Tabor⁶,⁷, Monica Tackett⁶, Marc Wegener⁶, Gao Wang¹⁰, Marsha M. Wheeler⁶, Amber Wright⁶ & Qian Yi⁶

⁶University of Washington, Seattle, WA, USA. ⁷Seattle Children’s Hospital, Seattle, WA, USA. ⁸University of Michigan, MI, USA. ⁹Hudson Alpha Institute of Technology, AL, USA. ¹⁰Baylor College of Medicine, TX, USA.