HHS Public Access

Author manuscript

*Genet Med*. Author manuscript; available in PMC 2022 January 29.

Published in final edited form as:

*Genet Med*. 2021 December ; 23(12): 2386–2393. doi:10.1038/s41436-021-01284-w.

# Large scale clinical exome sequencing uncovers the scope and severity of skin disorders associated with *MC1R* genetic variants

**Bryn S. Moore**, **Jonathan Z. Luo**, **Ann N. Stepanchick**, **Tooraj Mirshahi**

Department of Molecular and Functional Genomics, Geisinger, 100 North Academy Ave, Danville, PA 17837

## Abstract

**Purpose:** Genetic variation in *MC1R* is a main determinant of red hair color (RHC) phenotype which confers susceptibility to skin disorders.

**Methods:** We assessed the effects and function of *MC1R* variants identified in our clinical cohort of 135,947 participants with available exome sequencing using phenome-wide association scan (PheWAS). Expression and function of several variants was evaluated.

**Results:** We found 24 nonsense and 215 missense variants in *MC1R*. Many common missense *MC1R* variants are strongly associated with skin disorders including skin cancer; however, each variant shows different penetrance and expressivity. Severity of skin phenotype was well correlated with the magnitude of functional defect measured as receptor expression and α-MSH stimulated cAMP production. Remarkably, *MC1R* deletions and nonsense variants are only weakly associated with milder skin phenotypes.

**Conclusion:** Our comprehensive assessment of all *MC1R* variants in a large cohort clearly establish that individuals with some missense variants are more susceptible to severe skin disorders than those with *MC1R* deletions or nonsense variants.

## Introduction

Melanocortin 1 Receptor (MC1R), is a G-protein coupled receptor (GPCR), belonging to a family of 5 highly related melanocortin receptors[1, 2]. MC1R is expressed in the cutaneous melanocytes, located in the basal layer of the epithelium[3, 4]. Melanocytes synthesize



Correspondence should be addressed to: Tooraj Mirshahi Ph.D. tmirshahi@geisinger.edu.

**Author Contributions:** BSM performed experiments, analyzed data, prepared draft manuscript, co-wrote final manuscript; JZL performed PheWAS and SKAT-O analyses, prepared draft manuscript; ANS performed experiments, analyzed data; TM, conceptualized and supervised the study, co-wrote final manuscript.

Conceptualization: B.S.M., T.M.; Data curation: J.Z.L.; Formal Analysis: J.Z.L; Funding acquisition: T.M.; Investigation: B.S.M., A.N.S.; Writing – original draft: B.S.M.; Writing – review & editing: B.S.M., T.M.

Competing Interest: Authors declare no conflict of interest.

**Ethics Declaration:** This research was approved by the Geisinger Clinic Institutional Review Board and included 135,947 participants in the MyCode Health Initiative who have exome sequencing data obtained as part of the Geisinger-Regeneron DiscovEHR collaboration. All participants provided written informed consent, and all experiments were performed in accordance with relevant guidelines and regulations. The authors did not have access to any identifying information for the participants. The human phenotype and genotype data in this study were all deidentified by a "data broker" who was not involved in the study before any analysis was performed. De-identified clinical data was obtained from electronic health records (EHR).

two types of melanin: black/brown eumelanin and yellow/red pheomelanin, the balance and amount of which determine skin and hair color[5–8]. When stimulated by its agonist α-melanocyte-stimulating hormone (α-MSH) in response to UV radiation, MC1R stimulates the production of cAMP and results in the production of eumelanin, which has been shown to be photoprotective[9, 10].

MC1R consists of 317 amino acids forming 7 transmembrane domains, an extracellular N-terminus, and an intracellular C-terminus (Supplemental Figure 1). The C-terminus of MC1R is rather short and has a cysteine at position 315 that is palmitoylated[11]. Hundreds of protein-altering genetic variants in *MC1R* have been reported[12–14]. Some *MC1R* variants are associated with pale skin and red hair[15, 16] and an increased risk of melanoma and other skin cancers[17–21]. *MC1R* variants have been designated as low penetrance "r" or high penetrance "R" for red hair color (RHC)[22] and those with "R" designated variants are believed to be at higher risk of skin cancers[18, 23]. Among variants with MAF>0.005, *Val60Leu*, *Val92Met* and *Arg163Gln* are designated as "r" alleles and *Asp84Glu*, *Arg142His*, *Arg151Cys*, *Ile155Thr*, *Arg160Trp* and *Asp294His* are designated as "R" alleles showing variable penetrance[22], but *Ile155Thr* was later deemed a possible "r" allele[22, 24]. Individuals heterozygous or homozygous for "r" and "R" variants are all at an increased risk of cutaneous melanoma, non-melanoma skin cancers or actinic keratosis independent of pigmentation, showing variable expressivity[25–27]. Both "r" and "R" variants of *MC1R* show a spectrum of functional defects *in vitro* with either reduced cell surface expression and cAMP production, normal expression but reduced cAMP response, or normal to elevated cAMP production compared to wild-type MC1R. Furthermore, some *MC1R* variants were found to act in a dominant-negative manner by reducing cell surface receptor expression and intracellular cAMP signaling (Asp84Glu, Arg151Cys, Ile155Thr and Arg160Trp) or by only reducing cAMP signaling of co-expressed wild-type MC1R (Asp294His)[28].

Numerous studies have shown the association of *MC1R* variants with various cancers of the skin. However, many studies are limited by small case/control size[16, 26, 28], or unadjusted significance reporting (e.g., using $p<0.05$ without adjusting for multiple comparisons) for larger genome wide associations pooled analysis studies[23] or only examining a single clinical phenotype[18]. Even larger genome wide association studies have relatively small numbers (e.g., discovery population of 4336 control and 1650 cases and two replication cohorts of 964 case and 1149 control and 903 case and 1163 control)[29]. *MC1R* GWAS studies with a large number of participants have evaluated common genetic variant associations with different hair colors and do not evaluate for cancers of any type[30, 31]. Therefore, the spectrum of clinical phenotypes associated with *MC1R* variants remains largely unexplored.

Using exome sequencing data from the 135,947 participants of Geisinger-Regeneron DiscovEHR collaboration, we performed a phenome wide association scan (PheWAS) in a discovery cohort of 38,155 unrelated individuals and replicated these findings in a cohort of 51,712 unrelated individuals for whom we had an average of 14 years of longitudinal clinical data in a well-maintained electronic health record (EHR) system. The longitudinal data combined with WES allowed an unbiased approach using all phenotypes captured in

EHR mapped to 1866 PheCodes (see Methods). We determined the phenotypes associated with both missense (amino acids altering) and nonsense (predicted loss of function [pLOF] due to start-loss, early termination or frameshift) *MC1R* variants, as well as individuals with a copy number variant (CNV) having either one or three copies of *MC1R*. We found associations with missense *MC1R* variants and PheCodes only in the dermatologic and neoplasm categories which correlate with the levels of functional defect in each variant. Remarkably, we found that nonsense variants are only weakly associated with milder skin phenotypes.

## Materials and Methods:

### Study population, clinical variables, and whole exome sequencing (WES):

The research protocol was approved by the Geisinger Clinic Institutional Review Board and included 135,947 participants in the MyCode Health Initiative who have exome sequencing data obtained as part of the Geisinger-Regeneron DiscovEHR collaboration. Patients are consented to participate in MyCode and DiscovEHR from all clinics throughout the health system. All clinics share a uniform EHR that has been in place for over 20 years. Basic demographic information for participants in this study can be found in Supplemental Table 1. All participants provided written informed consent, and all experiments were performed in accordance with relevant guidelines and regulations. The authors did not have access to any identifying information for the participants. The human phenotype and genotype data in this study were all deidentified by a “data broker” who was not involved in the study before any analysis was performed. De-identified clinical data was obtained from EHR. Genomic DNA was isolated from patients’ blood or saliva. Whole exome sequencing was performed in collaboration with Regeneron Genetics Center as previously described[32]. Probes from NimbleGen (VCRome, referred to as VCR henceforth) or a modified version of the xGEN probe from Integrated DNA Technologies (IDT) were used for target sequence capture[33, 34]. Sequencing was performed by paired end 75bp reads on either an Illumina HiSeq2500 or NovaSeq. Coverage depth for all exome sites was sufficient to provide more than 20% coverage over 85% of the targeted bases in 96% of the VCR samples and 90% coverage for 99% of IDT samples. For *MC1R*, VCR samples had an average coverage of 30.6 at 93.5% of all sites in exon 3 (the coding exon) and IDT samples had an average coverage of 31.5 at 92.7% of all sites in exon 3. Alignments and variant calling were based on GRCh38 human genome reference sequence. Average read depth for sites with genetic variants (ref + alt) used in the analysis was 44.95 (range 39.8–52.8). The average allele balance (defined as alt / (ref + alt)) for variant sites was 49.8% (range 48.2–52.0). Nonsense or pLOF (predicted loss of function variants) are defined in this study as variants that cause a start-loss, frameshift or early termination/stop-gain of the encoded protein.

### Clinical traits, phenotype and PheCode definitions:

International Classification of Diseases Ninth (ICD-9) and Tenth (ICD-10) revision disease diagnosis codes were extracted from patients EHR. ICD codes were mapped to PheCodes using PheCodes Map 1.26 (https://phewascatalog.org/phecodes). For each individual, duplicate PheCode occurrences on the same date were dropped such that only one occurrence per date for a given PheCode remained. To ensure that individuals in the

study were adequately assessed for clinical history during clinical care, we restricted the analyses to individuals who were cases for at least one phenotype, which were defined as patients who were diagnosed with that phenotype on at least three distinct clinical encounters. Patients with zero diagnoses were deemed controls, whereas patients with one or two diagnoses were excluded from analysis of the phenotype i.e., they are neither case or control.

### PheWAS Analysis:

Phenome-Wide Association Scan (PheWAS) was performed to evaluate the effects of non-synonymous variants in *MC1R* with phenotypes encoded in EHR. First, second, and high-confidence third-degree relationships were removed using IBD estimates from Primus to obtain a maximal set of unrelated individuals as previously described[35]. Removing related subjects, resulted in a discovery cohort of 38,155 individuals (sequenced by VCR) and a replication cohort of 51,712 individuals (sequenced by IDT) in the final analyses. We used a threshold of at least 0.1% cases per code for each population (51 cases for IDT-sequenced and 38 for VCR-sequenced) to be included in the model. Associations were calculated using Firth logistic regression adjusted for age, sex, and ancestry using the first three principal components. Accounting for 10 principal components provided almost identical results. The analyses were performed assuming an additive genetic model; that is, we assume the risk due to an alternate allele is increased by r for heterozygotes and 2r for homozygotes. A circular plot of associations across the analyses for all variants was generated using Circos.

### SKAT-O:

For rare nonsense variants where single locus PheWAS was not feasible, we performed Optimized Sequence Kernel Association Test (SKAT-O) analysis[36]. SKAT-O was used to examine all nonsense variants (early terminations, frameshifts, and start-loss) except *Asn29LysfsTer14*, which was analyzed using PheWAS since there were sufficient number of subjects with this variant. Analysis was performed using the Robust SKAT-O method from SKAT version 2.0.0 in R. The analyses were performed assuming an additive genetic model. as described above.

### Molecular Biology:

Untagged human MC1R or N-terminal 3x HA tagged MC1R in pcDNA3.1+ were purchased from cDNA.org. Individual amino acid substitutions were made with the Quickchange site-directed mutagenesis kit (Stratagene). All constructs were confirmed by sequencing of the full-length clone.

### Cell Culture and transfection:

HEK293 cells (ATCC, Manassas, VA, USA) were cultured in MEM with 10% FBS at 37°C and 5% $CO_2$. For transient transfections, cells were transfected with plasmids described above by Xtremegene (Roche, Indianapolis, IN, USA) and used two days post-transfection. For the cAMP pGlo and ELISA assays, cells were transfected in one batch and then split for use in each assay. HEK293 cells stably expressing pGloSensor-20F cAMP plasmid (Promega) under Hygromycin selection were transfected with HA-MC1R, or HA-MC1R

harboring one of the variants, in wells of a 6-well dish. One day post-transfection, approximately 10,000 cells per well were added to white bottom (cAMP pGlo Assay) or clear poly-L-lysine coated (ELISA) 96 well dishes.

### cAMP pGlo Assay:

Two days post transfection, the media was carefully removed from the white-bottom plate and replaced with media containing 2% GloSensor cAMP reagent (Promega) and incubated at 37°C for 2 hours. Cells were stimulated with MC1R agonist α-MSH (0.01– 300 nM) or with 100 μM L-850851, a water soluble forskolin analog (to determine maximum cAMP), for 10 min and then the luminescence was read on a Spectramax M3 plate reader (Molecular Devices). Basal cAMP luminescence was subtracted, and cAMP values plotted as a percentage of the maximum cAMP measured for cells transfected with each *MC1R* variant. Data shown is from 3 independent experiments (mean ± SEM). Significant differences from wild-type were determined using one-way ANOVA with Dunnet's post-hoc.

### ELISA:

Two days post transfection, cells plated on clear poly L-lysine coated 96 well plates were washed with PBS and fixed with either methanol (for total expression) or 4% paraformaldehyde (for surface expression). Cells were then blocked with 1% milk and incubated in peroxidase conjugated anti-HA antibody. The plate was washed with TBS-T three times and then incubated with 100 μL 3,3′,5,5′-Tetramethylbenzidine Liquid Substrate (Sigma, St. Louis, MO, USA) for 30 minutes. 100 μL of 1 mol/L sulfuric acid was added to each well to stop the reaction. Absorbance was then read at 450 nm on a Spectramax M3 plate reader (Molecular Devices). The absorbance from untransfected cells was subtracted and then the cell surface labeled signal was plotted as a percentage of total signal (calculated as the non-permeabilized signal divided by the permeabilized signal ×100) and plotted. Total expression as a percentage of wild-type HA-MC1R for each experiment was also plotted. Data shown is from 3 independent experiments (mean ± SEM). Significant differences from wild-type were determined using one-way ANOVA with Dunnet's post-hoc.

## Results:

We have performed WES on 135,947 individuals using two different capture platforms: VCR and IDT. Patient DNA samples were collected from clinics within our integrated health system. However, all patient clinical data was captured in the same EHR. Among those with WES, we identified 89,867 unrelated individuals of which 38,155 were sequenced with the VCR platform and 51,712 were sequenced with the IDT platform. We used the 38,155 unrelated individuals (VCR) as discovery and the subsequent 51,712 unrelated individuals (IDT) as a replication cohort for this study.

In the discovery cohort we found 158 non-synonymous variants of *MC1R* consisting of 14 nonsense variants and 144 missense variants. In the replication cohort we found 199 non-synonymous variants of *MC1R*, consisting of 20 nonsense variants and 179 missense variants. In totality, the 239 non-synonymous variants of *MC1R* consisted of 24 nonsense and 215 missense variants (Supplemental Figure 1 and Supplemental

Table 2), of which 40 were novel variants (Supplemental Table 2)[12–14]. Ten variants (9 missense: *Val60Leu, Asp84Glu, Val92Met, Arg142His, Arg151Cys, Ile155Thr, Arg160Trp, Arg163Gln, Asp294His* and 1 nonsense: *Asn29LysfsTer14*) had MAF > 0.005 and MAF > 0.004 respectively, sufficient for single variant PheWAS analysis where the reference (wildtype) alleles were Ref, and each variant (separately) were the Risk allele.

PheWAS analysis showed strong association of *MC1R* variants to PheCodes related almost exclusively to dermatologic and neoplasm categories (Figure 1, Supplemental Figure 2, Supplemental Tables 3–5) in both the discovery and replication cohort. Figure 1 shows a Circos plot of all PheWAS data highlighting significant associations solely to the dermatological and neoplasm categories, except for one significant association between the *Ile155Thr* variant and an endocrine/metabolic PheCode. More granular associations between *MC1R* missense variants and dermatologic and neoplasm PheCodes can be seen in Figure 1 B and C. Odds ratios calculated for phenotypes identified in PheWAS are shown in the heat map in Figure 2 and Supplemental Tables 4 and 5. The strongest associations were found in the PheCode descriptions: melanomas of the skin, basal cell carcinoma, skin cancer, other non-epithelial cancer of skin, neoplasm of uncertain behavior of the skin, squamous cell carcinoma, actinic keratosis, fibrosis of skin and degenerative skin disorders and scar conditions. PheCode for degenerative skin disorders includes condition within ICD709.3 which encompasses a range of skin disorders including calcinosis, colloid milium, skin degeneration, skin deposits, senile dermatosis and subcutaneous calcification, it may also include patients with scars from previous procedures including skin cancer surgery. Of the 10 variants examined by PheWAS, only *Ile155Thr* was not significantly associated with any PheCodes in the discovery or replication cohorts (Figure 1 and 2, Supplemental Figure 2 and Supplemental Tables 4 and 5). *Arg151Cys* was significantly associated with all skin neoplasm and dermatologic PheCode descriptions listed above in both the discovery and replication cohorts (Figure 1 and 2, Supplemental Figure 2 and Supplemental Tables 4 and 5). All remaining variants tested were significantly associated with a number of PheCodes in both the discovery and replication cohorts.

To understand why some of the variants were significantly associated with these phenotypes and some were not, we decided to functionally assess each variant *in vitro*. Cells expressing MC1R, a $G_{\alpha s}$ coupled receptor, respond to α-MSH by producing cAMP. We evaluated *MC1R* variants for the ability to produce cAMP in response to α-MSH and for cell surface and total expression. Dose response curves for cAMP response to α-MSH showed a variety of variant affects: for example, Val92Met was not different from wildtype, whereas Asp84Glu had a right shifted dose response and a lower maximum response. Figure 3 A & B and Supplemental Figure 3 show data for all 9 amino acid substitutions tested. Calculated $EC_{50}$ for α–MSH showed that the $EC_{50}$s for Val60Leu, Asp84Glu, Arg142His, Arg151Cys, Ile155Thr and Asp294His were significantly different compared to wildtype MC1R (Figure 3A). α–MSH induced- maximum cAMP was significantly lower than wildtype for Asp84Glu, Arg142His, Arg151Cys, Ile155Thr, Arg160Trp, Arg163Gln and Asp294His (Figure 3B). Importantly, in cells from the same transfection as those used in the cAMP assay, the cell surface and total expression of all variants tested were similar to wildtype MC1R (Figure 3C and D). We plotted the odds ratios from the associated neoplasm and dermatologic phenotypes identified in PheWAS versus both the $EC_{50}$s and maximum

cAMP levels for the discovery (○) and replication cohorts (●) (Figures 4 and 5). Higher $EC_{50}$, i.e., reduced potency, corresponded to higher odds for these phenotypes (Figure 4 and 5 and Supplemental Figure 4), while lower maximum cAMP, i.e., reduced efficacy, corresponded to a higher odds ratio for these phenotypes (Figure 4 and 5 and Supplemental Figure 5).

In addition to the missense variants in *MC1R* found in the discovery and replication cohort, we found nonsense variants in *MC1R*. We recently showed that for MC1R family member melanocortin 4 receptor, truncation before the s-acylated Cys318 results in a non-functional receptor[37]. All nonsense *MC1R* variants identified in our cohort occur before the palmitoylation site[11], we therefore considered all of the nonsense variants true loss of function variants and included them in the analysis. *Asn29LysfsTer14* was the only nonsense variant where the number of heterozygous and homozygous individuals was sufficient for PheWAS analysis. *Asn29LysfsTer14* was significantly associated with actinic keratosis and marginally with skin cancer in both the discovery and replication cohorts (Figures 1 and 2, Supplemental Figure 2, Supplemental Tables 3–5). Since there was not a sufficient number of individuals with each of the other nonsense variant to perform PheWAS, we examined the remaining nonsense variants by robust SKAT-O, or optimized sequence kernel association test (Supplemental Table 6)[36]. Robust SKAT-O uses efficient resampling and saddle point approximation and aggregates the adjusted statistics to control for errors due to unbalanced case-control ratios. This analysis showed that nonsense variants were significantly associated with actinic keratosis (Supplemental Table 7) but not with neoplasms.

WES also revealed 21 individuals with copy number variants (CNV) of *MC1R*: 5 individuals with 1 copy of *MC1R* and 16 individuals with 3 copies of *MC1R*. Subjects with CNVs had few neoplasm or dermatologic phenotypes in their EHR (Supplemental Table 7). Interestingly, one of the subjects with an *MC1R* CNV deletion had an *Arg151Cys* variant in the remaining copy. This individual had neoplasm of uncertain behavior of skin. Seven subjects with *MC1R* duplication also harbored other variants in *MC1R*: 5 with *Arg160Trp*, 1 with *Arg163Gln* and 1 with *Met128Thr*. Only one of the individuals carrying an *MC1R* duplication and an *Arg160Trp* variant had a neoplasm phenotype.

## Discussion:

We have conducted, to our knowledge, the largest and most comprehensive study of *MC1R* genotype/phenotype relationship to date. We used exome sequencing and longitudinal clinical data from two cohorts with a total of >135,000 participants, with almost 90,000 unrelated participants used for association analyses, combined with *in vitro* assessment for function and expression of variants found in the sequencing data to conclusively establish the scope and spectrum of effects for common *MC1R* variants in skin disorders and neoplasms. We provide novel data on loss of function and copy number deletion variants to show a lack of strong association with severe phenotypes, as well as strong association of many relatively common missense variants with skin disorders and neoplasms. These data establish that individuals with missense *MC1R* variants that impair receptor function are at highest risk for the more severe neoplasms associated with skin.

Knowledge of *MC1R* genetic effects on diseases of the skin have primarily relied on case/control studies with relatively small number of individuals. Additionally, family studies could artificially enrich the influence of a particular genetic makeup on a phenotype. GWAS have identified some *MC1R* variants associated with melanoma[29] but are not well suited to determine associations with a wide spectrum of phenotypes. *MC1R* variants have previously been designated as low penetrance "r" or high penetrance "R" for red hair color with many studies grouping variants based on the R/r designation in their analyses. In this study we looked at the association of clinical traits with individual *MC1R* variants, not with R/r classifications of variants. Grouping variants whose encoded proteins function differently would not accurately capture the individual differences among these variants and their associated clinical phenotypes even if they have all been associated with the red hair color phenotype. Additionally, MC1R related skin cancers are independent from hair color[26, 27]. For example, each of the missense variants maintain some function while early frameshifts and terminations do not. Grouping variants into "R" and "r" groups does not allow for the nuanced associations we found when we examined these variants individually. Even among missense variants with similar cAMP responses and expression, *Ile155Thr* and *Asp294His*, we found significant differences in associations of each variant with phenotypes. Rare, loss of function variants however can be grouped together for analysis because they fail to express or function.

We established discovery and replication cohorts of 38,155 and 51,712 unrelated individuals for these analyses, eliminating the inherent bias that can occur in family studies. These individuals were not selected based on RHC phenotype or history of skin cancer as in other studies[13–15]. In totality we found 239 non-synonymous variants in *MC1R* consisting of 24 nonsense and 215 missense variants including 40 previously unreported variants[38–40] (Supplemental Figure 1 and Supplemental Table 2). Ten variants (9 missense: *Val60Leu*, *Asp84Glu*, *Val92Met*, *Arg142His*, *Arg151Cys*, *Ile155Thr*, *Arg160Trp*, *Arg163Gln*, *Asp294His* and 1 nonsense: *Asn29LysfsTer14*) had MAF > 0.004, sufficient for PheWas analysis. The remaining nonsense variants were grouped and evaluated by SKAT-O analysis due to the small number of heterozygotes for each variant.

In the discovery and replication cohorts *MC1R* variants *Val60Leu*, *Asp84Glu*, *Val92Met*, *Arg142His*, *Arg151Cys*, *Arg160Trp*, *Arg163Gln* and *Asp294His* were significantly associated with actinic keratosis and skin cancer (Odds ratios>1 and p values $<4.46 \times 10^{-5}$ - $6.31 \times 10^{-5}$). Other non-epithelial cancer of the skin and degenerative skin conditions and other dermatoses were significantly associated with the variants *Asp84Glu*, *Val92Met*, *Arg142His*, *Arg151Cys*, *Arg160Trp*, *Arg163Gln* and *Asp294His*. *Ile155Thr* has been categorized as an "R" allele and then later as a possible "r" allele[22, 24], but regardless of the categorization *Ile155Thr* has reportedly been associated with melanomas[17, 23]. Interestingly *Ile155Thr* was not significantly associated with any skin cancer or dermatologic condition in either cohort, contradicting earlier reports. In addition to skin cancer, other non-epithelial cancer of the skin, actinic keratosis, and degenerative skin conditions, those having the *Arg151Cys* variant were significantly associated with neoplasm of uncertain behavior of skin, melanomas of skin, squamous cell carcinoma and basal cell carcinoma.

Functionally there were a variety of effects observed for these missense variants from almost normal functionality (Val92Met) to significantly impaired (significantly shifted $EC_{50}$ and significantly reduced maximum cAMP) (Asp84Glu, Arg142His, Arg151Cys, Ile155Thr and Asp294His). Interestingly, Ile155Thr significantly altered both the $EC_{50}$ and maximum cAMP. When we examined odds ratios versus the $EC_{50}$ or maximum cAMP for skin cancer phenotypes particularly for skin cancer or melanomas we observed that higher odds ratio correlated well with higher $EC_{50}$ and lower maximum cAMP response.

The *Asn29LysfsTer14* variant was significantly associated with actinic keratosis and skin cancer in both the discovery and replication cohorts, though the association with skin cancer was only marginally significant (Figure 1B). SKAT-O analysis revealed that the other *MC1R* loss of function variants were not at increased risk of any skin cancers or diseases of the skin aside from actinic keratosis. Additionally, only two of five individuals with *MC1R* CNV deletion had neoplasm in their EHR, but one also had an *Arg151Cys* variant in the remaining copy which is most likely driving the phenotype given the strong association of *Arg151Cys* variant with multiple skin and neoplasm phenotypes (Supplemental Table 8). Combined data from protein truncating and copy number variants strongly suggest that a single functioning copy of *MC1R* is sufficient to protect from the more severe skin disorders associated with the *MC1R* missense variants that impair receptor function.

We have combined genetic data from a very large cohort with a phenotype agnostic approach to establish the scope and spectrum of *MC1R* genotype/phenotype relationship. We could not replicate previously reported association of *Ile155Thr* with various skin phenotypes, while all other missense variants with MAF>0.005 were strongly associated with multiple skin and neoplasm phenotypes. Most importantly, we had 1781 individuals with total loss of function variants (early terminations, frameshifts) and 5 individuals with copy number deletion of *MC1R* due to large chromosomal deletions. These individuals effectively only have one copy of *MC1R* but showed surprisingly weak associations with skin phenotypes and almost no associations with neoplasms. Additionally, one individual who had a CNV-deletion combined with a missense variant of *MC1R* had neoplasm of uncertain behavior of skin.

These data strongly suggest that a single functional copy of *MC1R*, in the absence of a missense variant in the other copy, is necessary and sufficient to produce enough cAMP to be photoprotective due to production of eumelanin. Additionally, heterozygous, and homozygous individuals for missense variants *Asp84Glu*, *Val92Met*, *Arg142His*, *Arg151Cys*, *Arg160Trp*, *Arg163Gln*, *Asp294His* and to some extent *Val60Leu* are at a greater risk of skin neoplasms and other dermatoses than reference, but those with nonsense or CNV variants of *MC1R* are not. Our findings provide new generalizable guidelines for use of *MC1R* genetics in assessing risk of skin disorders, including skin cancer, independently of the red hair phenotype.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments:

This work was supported by GM111913 to T.M.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Data Availability:

The data supporting the findings of this study are available within the article and its Supplementary Data files. Additional information for reproducing the results described in the article is available upon reasonable request and subject to a data use agreement.

## References

1. Cai M & Hruby VJ The Melanocortin Receptor System: A Target for Multiple Degenerative Diseases. Curr. Protein Pept. Sci 17, 488–496 (2016). [PubMed: 26916163]
2. Moscowitz AE et al. The Importance of Melanocortin Receptors and Their Agonists in Pulmonary Disease. Front. Med. (Lausanne) 6, 145 (2019). [PubMed: 31316990]
3. Mountjoy KG, Robbins LS, Mortrud MT & Cone RD The cloning of a family of genes that encode the melanocortin receptors. Science 257, 1248–1251 (1992). [PubMed: 1325670]
4. Roberts DW, Newton RA, Beaumont KA, Helen Leonard J & Sturm RA Quantitative analysis of MC1R gene expression in human skin cell cultures. Pigment Cell Res 19, 76–89 (2006). [PubMed: 16420249]
5. Herraiz C, Garcia-Borron JC, Jimenez-Cervantes C & Olivares C MC1R signaling. Intracellular partners and pathophysiological implications. Biochim. Biophys. Acta Mol. Basis Dis 1863, 2448–2461 (2017). [PubMed: 28259754]
6. Thong HY, Jee SH, Sun CC & Boissy RE The patterns of melanosome distribution in keratinocytes of human skin as one determining factor of skin colour. Br. J. Dermatol 149, 498–505 (2003). [PubMed: 14510981]
7. Matts PJ, Dykes PJ & Marks R The distribution of melanin in skin determined in vivo. Br. J. Dermatol 156, 620–628 (2007). [PubMed: 17493065]
8. d'Ischia M et al. Melanins and melanogenesis: from pigment cells to human health and technological applications. Pigment Cell. Melanoma Res 28, 520–544 (2015). [PubMed: 26176788]
9. Duffy DL et al. Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. Hum. Mol. Genet 13, 447–461 (2004). [PubMed: 14709592]
10. Rouzaud F, Kadekaro AL, Abdel-Malek ZA & Hearing VJ MC1R and the response of melanocytes to ultraviolet radiation. Mutat. Res 571, 133–152 (2005). [PubMed: 15748644]
11. Chen S et al. Palmitoylation-dependent activation of MC1R prevents melanomagenesis. Nature 549, 399–403 (2017). [PubMed: 28869973]
12. Tate JG et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 47, D941–D947 (2019). [PubMed: 30371878]
13. Stenson PD et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum. Genet 136, 665–677 (2017). [PubMed: 28349240]
14. McLaren W et al. The Ensembl Variant Effect Predictor. Genome Biol 17, 122–4 (2016). [PubMed: 27268795]
15. Valverde P, Healy E, Jackson I, Rees JL & Thody AJ Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat. Genet 11, 328–330 (1995). [PubMed: 7581459]
16. Box NF, Wyeth JR, O'Gorman LE, Martin NG & Sturm RA Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair. Hum. Mol. Genet 6, 1891–1897 (1997). [PubMed: 9302268]

17. Scherer D et al. MC1R variants associated susceptibility to basal cell carcinoma of skin: interaction with host factors and XRCC3 polymorphism. Int. J. Cancer 122, 1787–1793 (2008). [PubMed: 18067130]
18. Han J, Kraft P, Colditz GA, Wong J & Hunter DJ Melanocortin 1 receptor variants and skin cancer risk. Int. J. Cancer 119, 1976–1984 (2006). [PubMed: 16721784]
19. Kanetsky PA et al. Population-based study of natural variation in the melanocortin-1 receptor gene and melanoma. Cancer Res 66, 9330–9337 (2006). [PubMed: 16982779]
20. Raimondi S et al. MC1R variants, melanoma and red hair color phenotype: a meta-analysis. Int. J. Cancer 122, 2753–2760 (2008). [PubMed: 18366057]
21. Williams PF, Olsen CM, Hayward NK & Whiteman DC Melanocortin 1 receptor and risk of cutaneous melanoma: a meta-analysis and estimates of population burden. Int. J. Cancer 129, 1730–1740 (2011). [PubMed: 21128237]
22. Beaumont KA, Liu YY & Sturm RA The melanocortin-1 receptor gene polymorphism and association with human skin cancer. Prog. Mol. Biol. Transl. Sci 88, 85–153 (2009). [PubMed: 20374726]
23. Pasquali E et al. MC1R variants increased the risk of sporadic cutaneous melanoma in darker-pigmented Caucasians: a pooled-analysis from the M-SKIP project. Int. J. Cancer 136, 618–631 (2015). [PubMed: 24917043]
24. Beaumont KA et al. Melanocortin MC(1) receptor in human genetics and model systems. Eur. J. Pharmacol 660, 103–110 (2011). [PubMed: 21199646]
25. Duffy DL et al. Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. J. Invest. Dermatol 130, 520–528 (2010). [PubMed: 19710684]
26. Bastiaens MT et al. Melanocortin-1 receptor gene variants determine the risk of nonmelanoma skin cancer independently of fair skin and red hair. Am. J. Hum. Genet 68, 884–894 (2001). [PubMed: 11254446]
27. Jacobs LC et al. IRF4, MC1R and TYR genes are risk factors for actinic keratosis independent of skin color. Hum. Mol. Genet 24, 3296–3303 (2015). [PubMed: 25724930]
28. Beaumont KA et al. Receptor function, dominant negative activity and phenotype correlations for MC1R variant alleles. Hum. Mol. Genet 16, 2249–2260 (2007). [PubMed: 17616515]
29. Bishop DT et al. Genome-wide association study identifies three loci associated with melanoma risk. Nat. Genet 41, 920–925 (2009). [PubMed: 19578364]
30. Morgan MD et al. Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. Nat. Commun 9, 5271–z (2018). [PubMed: 30531825]
31. Hysi PG et al. Publisher Correction: Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. Nat. Genet 51, 1190–3 (2019).
32. Carey DJ et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. Genet. Med 18, 906–913 (2016). [PubMed: 26866580]
33. Dewey FE et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 354, 10.1126/science.aaf6814 (2016).
34. Van Hout CV et al. Exome sequencing and characterization of 49,960 individuals in UK Biobank. . Nature (2020).
35. Staples J et al. PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. Am. J. Hum. Genet 95, 553–564 (2014). [PubMed: 25439724]
36. Zhao Z et al. UK Biobank Whole-Exome Sequence Binary Phenome Analysis with Robust Region-Based Rare-Variant Test. Am. J. Hum. Genet 106, 3–12 (2020). [PubMed: 31866045]
37. Moore BS & Mirshahi T Genetic variants help define the role of the MC4R C-terminus in signaling and cell surface stability. Sci. Rep 8, 10397–3 (2018). [PubMed: 29991773]
38. Ringholm A et al. Pharmacological characterization of loss of function mutations of the human melanocortin 1 receptor that are associated with red hair. J. Invest. Dermatol 123, 917–923 (2004). [PubMed: 15482480]

39. Newton RA et al. Activation of the cAMP pathway by variant human MC1R alleles expressed in HEK and in melanoma cells. Peptides 26, 1818–1824 (2005). [PubMed: 15992961]

40. Scott MC et al. Human melanocortin 1 receptor variants, receptor function and melanocyte response to UV radiation. J. Cell. Sci 115, 2349–2355 (2002). [PubMed: 12006619]

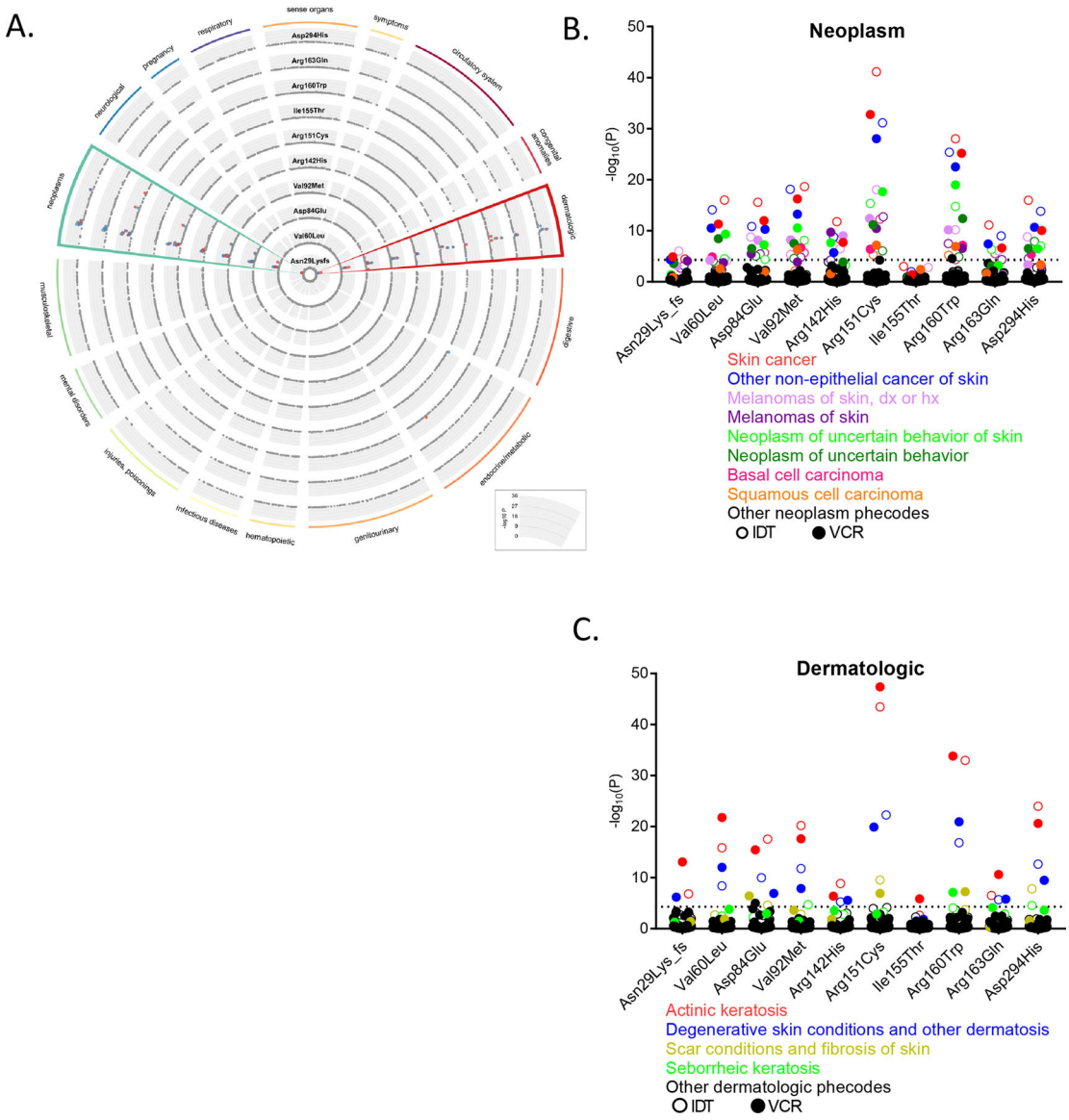


**Figure 1:**

PheWAS for 10 common variants of *MC1R*, 9 missense variants and 1 nonsense *MC1R* variant. **A.** Circle Graph showing associations for 1886 PheCodes mapped from EHR. The dotted line represents the $-\log_{10}$(Bonferroni corrected p value) for each variant (range is 4.2–4.28 for IDT sequenced group and 4.29–4.55). Significant association were almost exclusively observed among dermatologic and neoplasm PheCode classes in the PheWAS analysis. **B-C.** Manhattan plot for association of dermatologic and neoplasms of skin PheCodes with *MC1R* variants.

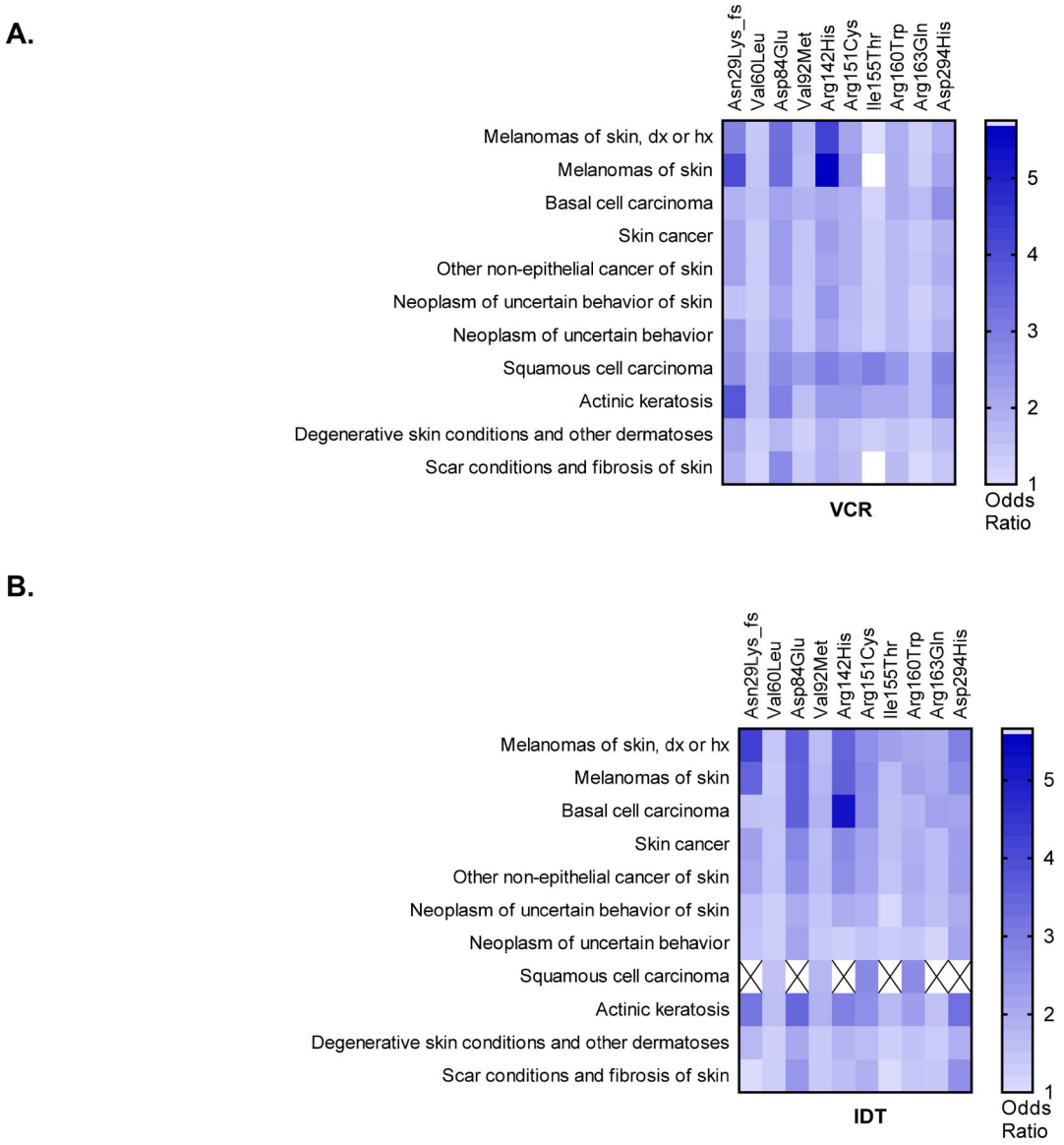


**Figure 2:**

Heatmap of Odds Ratios from PheWAS of common *MC1R* variants shows the association of *MC1R* common variants with neoplasms of skin or dermatologic phenotypes. A) PheWAS-VCR, B). PheWAS-IDT. X = minimum number of observations not met (38 for VCR and 51 for IDT).

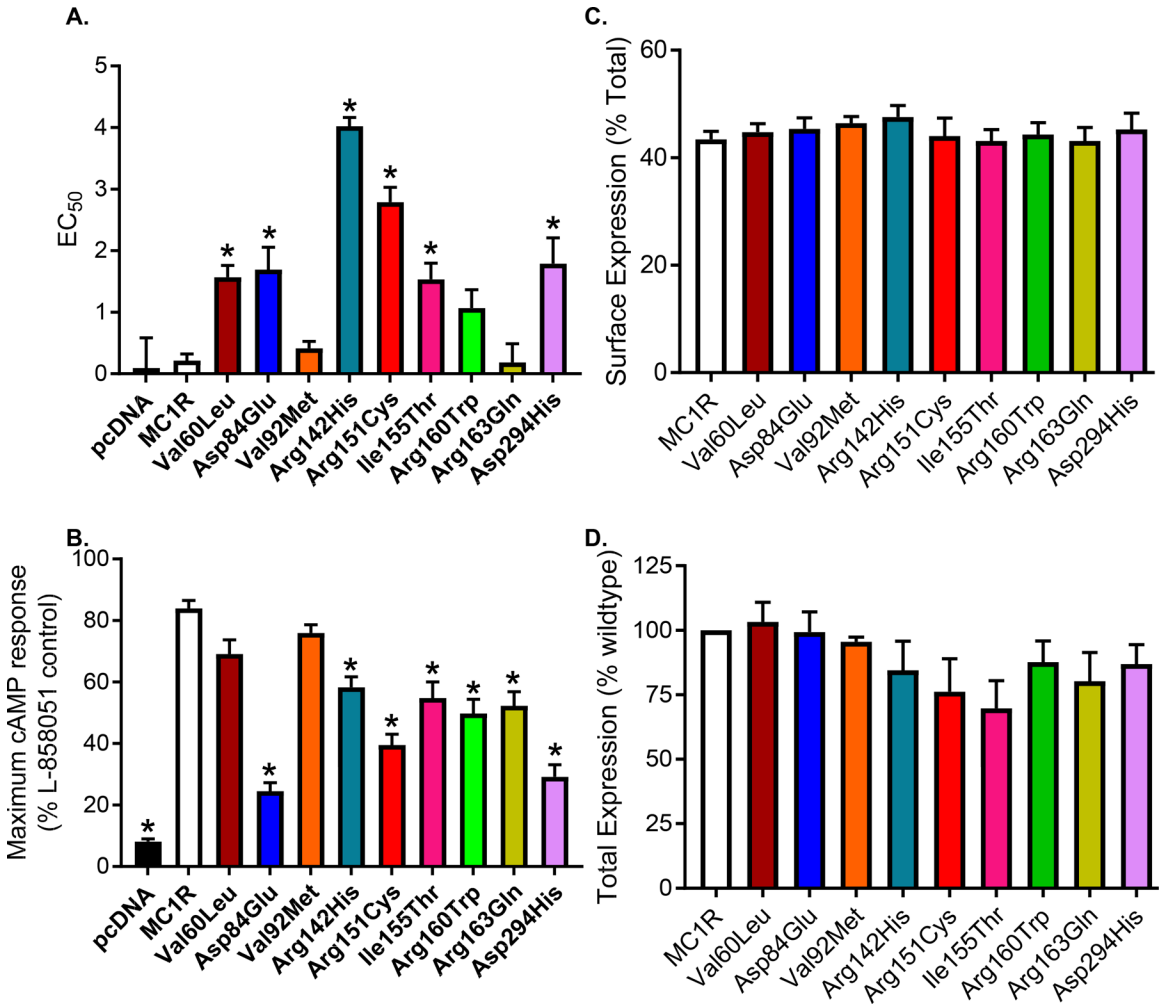


**Figure 3:**

In vitro functional data for missense variants with MAF > 0.005. **A.** Calculated $EC_{50}$ for α –MSH for missense variants. The $EC_{50}$s for Val60Leu, Asp84Glu, Arg142His, Arg151Cys, Ile155Thr and Asp294His variants are significantly different compared to wildtype MC1R. **B.** Calculated α –MSH induced maximum cAMP (% of L-858051) for missense variants. The maximum cAMP for Asp84Glu, Arg142His, Arg151Cys, Ile155Thr, Arg160Trp, Arg163Gln and Asp294His variants are significantly lower than wildtype MC1R. **C.** Cell surface and **D.** total expression of common missense variants are similar to wildtype MC1R.

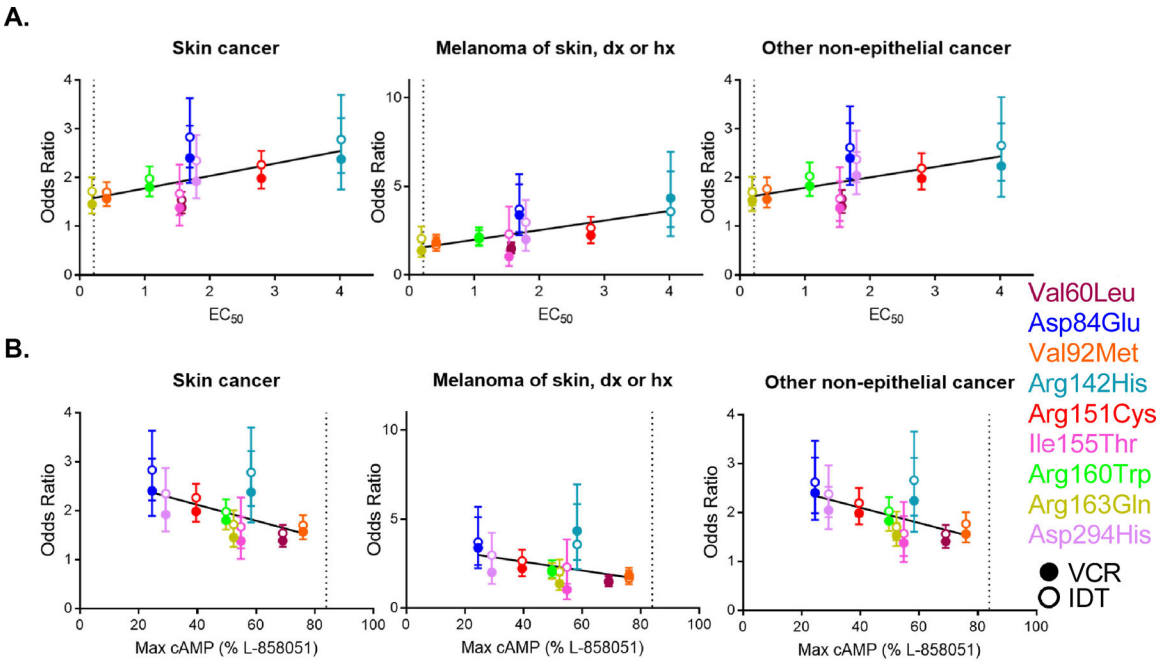


**Figure 4:**

Relationship of neoplasms and functional consequences of common *MC1R* missense variants. **A.** Plots of the odds ratio for various skin neoplasms versus $EC_{50}$ for each common variant. The dotted line represents the $EC_{50}$ for wildtype MC1R. **B.** Plots of the odds ratio for various skin neoplasms versus maximum α-MSH induced cAMP for each common variant. Error bars represent 95% CI. The dotted line represents wildtype MC1R max cAMP (%L-858051). Symbols are color coded to match bars in Figure 3.

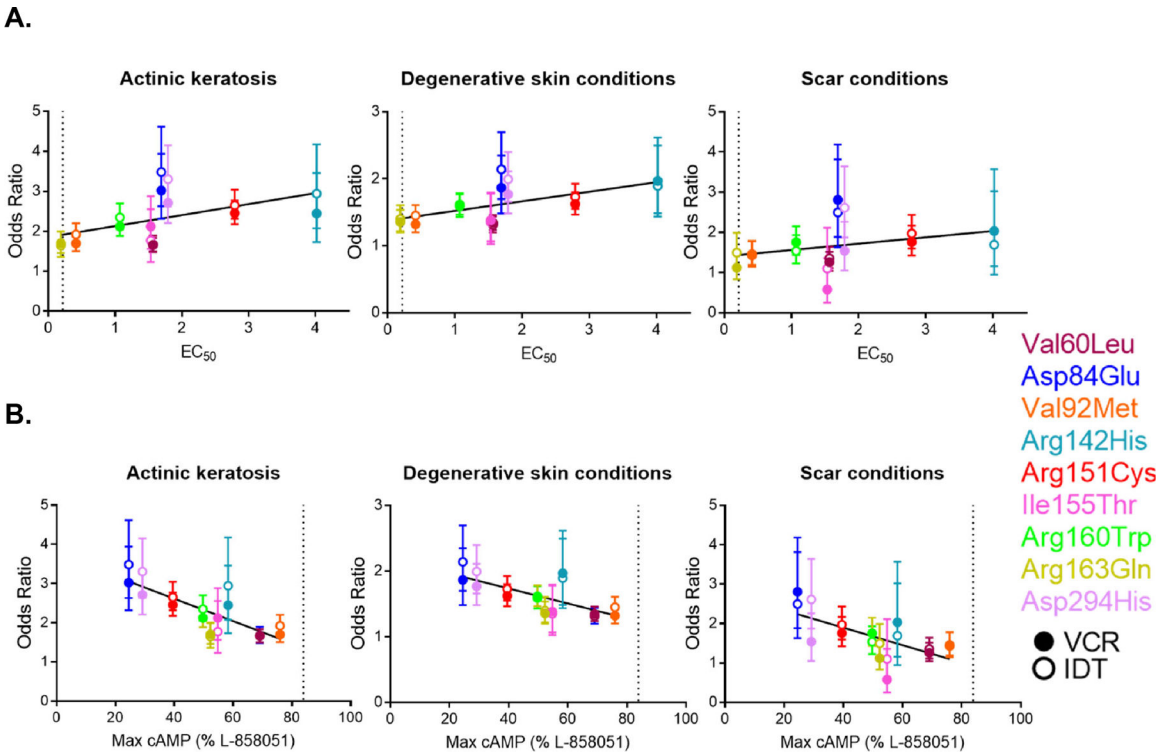


**Figure 5:**

Relationship of dermatologic clinical traits and functional consequences of common *MC1R* missense variants. **A.** Plots of the odds ratio for dermatologic phenotypes versus $EC_{50}$ for each common variant. The dotted line represents the $EC_{50}$ for wildtype MC1R. **B.** Plots of the odds ratio for dermatologic phenotypes versus maximum α-MSH induced cAMP for each common variant. Error bars represent 95% CI. The dotted line represents wildtype MC1R max cAMP (%L-858051). Symbols are color coded to match bars in Figure 3.