Combining *genetic* and exposure *data* significantly improves *risk prediction* for *skin cancer*

Pierre Fontanillas, Babak Alipanahi, Michaela Johnson, Catherine Wilson, 23andMe Research Team, Steve Pitts, Robert Gentleman, Adam Auton.

10/17/2018

https://research.23andme.com/
• **Basal cell carcinoma** (BCC) and **squamous cell carcinoma** (SCC) are the more common forms

• **Melanoma** is generally better studied due to its more aggressive nature
• **Basal cell carcinoma** (BCC) and **squamous cell carcinoma** (SCC) are the more common forms
• **Melanoma** is generally better studied due to its more aggressive nature

• Epidemiological studies had identified main **risk factors** (UV exposure and pigmentation type)
• However, studies were generally **small**, tested a **restricted set of risk factors**, and did not use **out-of-sample validation**
• Difficult to **build good predictive models**
Replicate/identify risk factors for BCC, SCC, and melanoma

• In a **large cohort**
• **Exposure** risk factors
• **Genetic risk score** (GRS)

**Improve skin cancer prediction:**

• Build the **best predictive models**
• [[ GxE interaction: GRS x Exposure ]]
• **Out-of-sample** validation
5M+ customers

~80% consent to research

75% ≥1 survey

1.5 B+ survey questions answered

23andMe cohort

1.6M genotyped variants

57M imputed variants
New cancer survey containing 72 questions:

### Personal history of cancer
- Type of cancer
- Age of diagnosis
- Body location
- Diagnosis type
- Treatments

### Risk factors
- Eye, hair, and skin colors
- Freckles
- Moles
- Skin response to sun exposure
- Sunbathing frequency
- Typical weekly sun exposure

### Family history of cancer
- Cancer history of relatives (father, mother, siblings, and children)

---


Goal

Questions

Design

Training set

Validation set

Conclusions

**Number of participants**

<table>
<thead>
<tr>
<th>Period</th>
<th>European ancestry</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 2016</td>
<td>23andMe participants</td>
<td>88,985</td>
</tr>
<tr>
<td>Jul 2016</td>
<td>23andMe participants</td>
<td>103,008</td>
</tr>
</tbody>
</table>

**Training Set**

- BCC: 14,898
- Melanoma: 3,998
- SCC: 7,479

- 2171
- 758
- 1643
- 4541
- 2754
- 942
- 9472
New cancer survey

Personal history of cancer
34 questions
- Type of cancer
- Age of diagnosis
- Body location
- Diagnosis type
- Treatment

Risk factors
23 questions
- Eye, hair, and skin colors
- Freckles
- Moles
- Skin response to sun exposure
- Sunbathing frequency
- Typical weekly sun exposure

Family history of cancer
16 questions
- Cancer history of relatives (father, mother, siblings, and children)

Generalized linear models:

BCC
SCC  ~ Skin color + Large moles + Sunbathing frequency + BMI + …

Melanoma
(cases / controls)
### New cancer survey

<table>
<thead>
<tr>
<th>Personal history of cancer</th>
<th>34 questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Type of cancer</td>
<td></td>
</tr>
<tr>
<td>- Age of diagnosis</td>
<td></td>
</tr>
<tr>
<td>- Body location</td>
<td></td>
</tr>
<tr>
<td>- Diagnosis type</td>
<td></td>
</tr>
<tr>
<td>- Treatment</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>23 questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Eye, hair, and skin colors</td>
<td></td>
</tr>
<tr>
<td>- Freckles</td>
<td></td>
</tr>
<tr>
<td>- Moles</td>
<td></td>
</tr>
<tr>
<td>- Skin response to sun exposure</td>
<td></td>
</tr>
<tr>
<td>- Sunbathing frequency</td>
<td></td>
</tr>
<tr>
<td>- Typical weekly sun exposure</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family history of cancer</th>
<th>16 questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Cancer history of relatives (father, mother, siblings, and children)</td>
<td></td>
</tr>
</tbody>
</table>

### Generalized linear models:

**BCC**  
**SCC**  
**Melanoma**

~ Demographic + Genetic Risk Score + Family History of skin cancer + Moles + Susceptibility + Exposure + Misc.

<table>
<thead>
<tr>
<th>Best Model: 32 factors</th>
<th>Age 1 factor</th>
<th>Sex 1 factor</th>
<th>Ancestry 4 factors</th>
<th>Training set 8 factors</th>
<th>Validation set 8 factors</th>
<th>Conclusions 7 factors</th>
<th>Design</th>
</tr>
</thead>
</table>
Number of variants in GRS:
- **BCC**: 47
- **SCC**: 14
- **Melanoma**: 18

Total deviance explained by best models (32 factors):
- Melanoma: 19.9%
- SCC: 20.0%
- BCC: 21.8%
Dysplastic moles
Actinic keratosis before age of 40
Presence of large moles
Moles in your right arm
Dysplastic moles
Actinic keratosis before age of 40
Presence of large moles
Moles in your right arm

Sunbathing frequency before age of 30
Adulthood latitude
Physical activity
Typical sun exposure per week
Childhood latitude
Tanning bed
Outdoor job
Adulthood elevation

Skin, hair, and eye colors
Freckling tendency
Blisters caused by sunburns
Skin response to 3-4 hours of sun exposure
Sun hair lightening
Skin response to sun exposure without sun protection

BMI
Alcohol consumption
Weight
Smoking
Morning person
Desk clean person
Seasonal allergies
Presence of large moles

Skin color

Sunbathing frequency before age of 30

BMI

Melanoma

SCC

BCC
**Goal**

**Questions**

**Design**

**Training set**

**Validation set**

**Conclusions**

**Melanoma:** 25 predictive models reviewed; AUC = 0.755

% of high risk individuals with >2 fold prevalence increase (relative to mean prevalence)

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>BCC</th>
<th>SCC</th>
<th>Melanoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRS</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>GRS + Exposure</td>
<td>5%</td>
<td>4%</td>
<td>1%</td>
</tr>
<tr>
<td>Demographics (age + sex + ancestry)</td>
<td>6%</td>
<td>21%</td>
<td>12%</td>
</tr>
<tr>
<td>Total</td>
<td>25%</td>
<td>31%</td>
<td>32%</td>
</tr>
</tbody>
</table>
Age of diagnosis

- BCC: 4.9 years
- SCC: 7.2 years
- Melanoma: 7.4 years
- Validation set: 7.5 years
• New cancer survey + 100K samples training set
• 32 risk factors in predictive models
• Out-of-sample validation (90K samples)
• **New cancer survey** + **100K samples** training set
• **32 risk factors** in predictive models
• **Out-of-sample** validation (**90K** samples)

• BCC, SCC and melanoma have **similar risk factors**,  
• Melanoma **quantitatively different**
• New cancer survey + 100K samples training set
• 32 risk factors in predictive models
• Out-of-sample validation (90K samples)

• BCC, SCC and melanoma have similar risk factors,
• Melanoma quantitatively different

• Genetic and exposure risk scores identified individuals at high risk for skin cancer
• [[ Absence of large GxE (no evidence for GRS x Exposure) ]]
Acknowledgements

23andMe research participants

23andMe Research Team:

Michelle Agee
Babak Alipanahi
Adam Auton
Robert K. Bell
Katarzyna Bryc
Sarah L. Elson
Pierre Fontanillas
Nicholas A. Furlotte
David A. Hinds
Bethann S. Hromatka
Yunru (Claire) Huang
Karen E. Huber
Aaron Kleinman
Nadia K. Litterman
Matthew H. McIntyre
Joanna L. Mountain
Carrie A.M. Northover
Steven J. Pitts
Cody Sam
J. Fah Sathirapongsasuti
Olga V. Sazonova
Janie F. Shelton
Suyash Shringarpure
Chao Tian
Joyce Y. Tung
Vladimir Vacic
Catherine H. Wilson

23andMe employees

https://research.23andme.com/