NISTIR XXXX Draft

Ongoing Face Recognition Vendor Test (FRVT)
Part 1: Verification

Patrick Grother
Mei Ngan
Information Access Division
Information Technology Laboratory

This publication is available free of charge from:
https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing

National Institute of Standards and Technology
U.S. Department of Commerce

DISCLAIMER

ABOUT THIS REPORT

This report documents the verification-track of the ongoing Face Recognition Vendor Test. The report will be updated continuously as new algorithms are evaluated, as new datasets are added, and as new analyses are included. Comments and suggestions should be directed to frvt@nist.gov.

Contents

List of Tables

List of Figures

1 Algorithm summaries

Developer	Short	Seq. number	Validation date	Size config data (KB)	Template size (B)	Template time (ms)	GPU
Dermalog	dermalog	001	2017-02-22	0	1041 ± 47	182 ± 11	No
Dermalog	dermalog	002	2017-02-22	0	1041 ± 47	139 ± 13	No
N-Tech Lab	ntech	000	2017-03-13	191530	2906 ± 1	550 ± 28	No

Table 1: The entries show the algorithms included in this report

2 Datasets

2.1 Visa images

- ▷ The number of images is O(10^5)
- ▷ The number of subjects is O(10^5)
- ▷ The number of subjects with two images O(10^4)
- ▷ The images have geometry in reasonable conformance with the ISO/IEC 19794-5 Full Frontal image type.
- ▷ The images are of size 252x300 pixels. The mean IOD is 69 pixels.
- ▷ The images are of subjects from greater than 100 countries, with significant imbalance due to visa issuance patterns.
- ▷ The images are of subjects of all ages, including children, again with imbalance due to visa issuance demand.

3 Results

3.1 Overall accuracy

Goals:

- ▷ To state overall accuracy.
- ▷ To compare algorithms.

Method:

- ▷ The images are from the visa dataset.
- ▷ The comparisons are of visa photos against visa photos.
- ▷ The number of genuine comparisons is O(10^4).
- ▷ The number of impostor comparisons is O(10^{10}).
- ▷ The comparisons are fully zero-effort, meaning impostors are paired without attention to sex, age or other covariates.
- ▷ The number of persons is O(10^5).
- ▷ The number of images used to make a template is 1.
- ▷ The number of templates used to make each comparison score is two corresponding to simple one-to-one verification.

Summary: Core algorithm accuracy is stated via the error tradeoff characteristics of Figure 1.

Figure 2 shows dependence of false match rate on algorithm score threshold. This allows a deployer to set a threshold to target a particular false match rate appropriate to the security objectives of the application. Note that false match rates vary with the ethnicity, age, and sex, of the enrollee and impostor - see section 3.3.

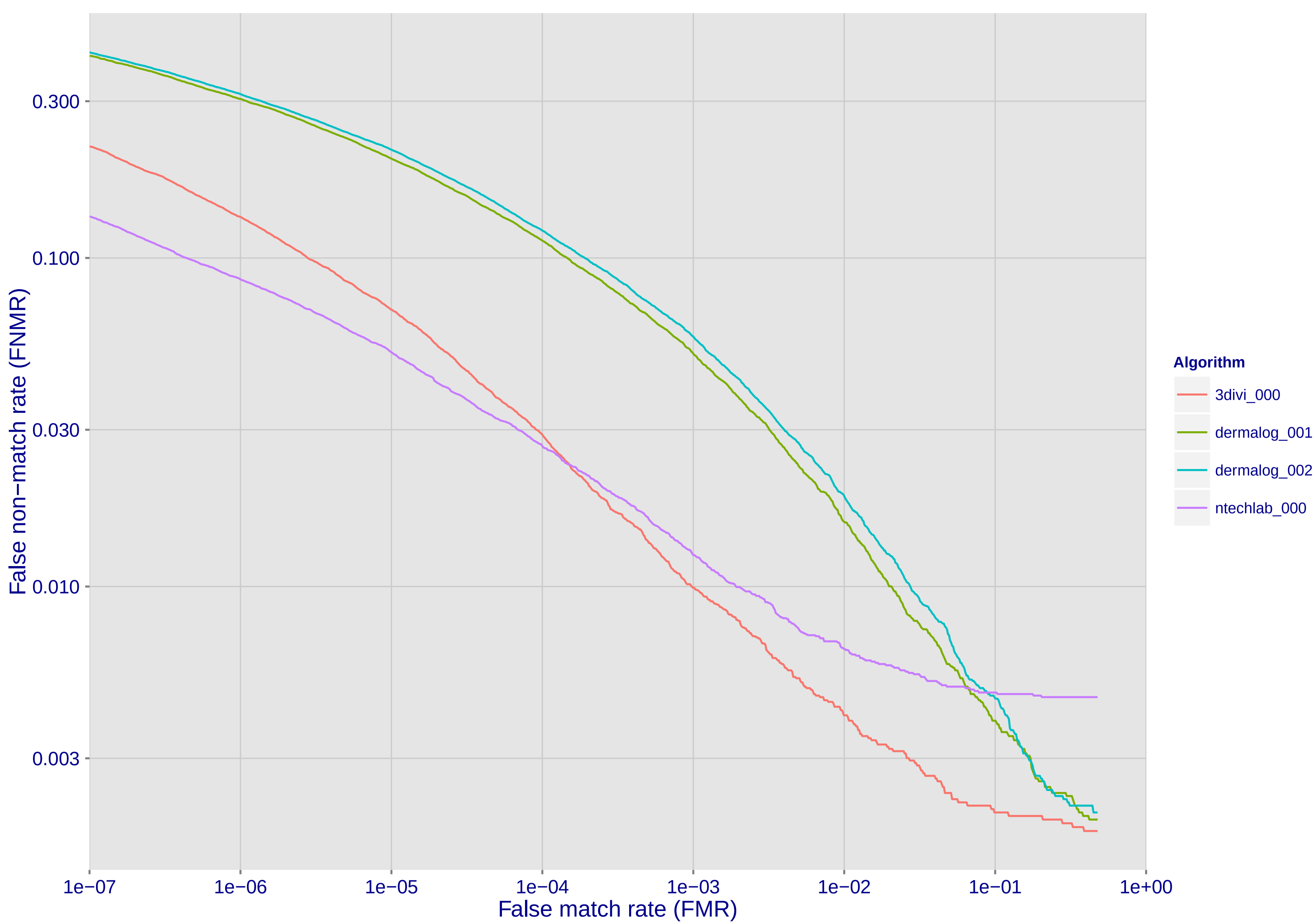

*Figure 1: For the visa images, detection error tradeoff (*DET*) characteristics showing false non-match rate vs. false match rate plotted paratemetrically on threshold, T. The scales are logarithmic in order to show many decades of* FMR.

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

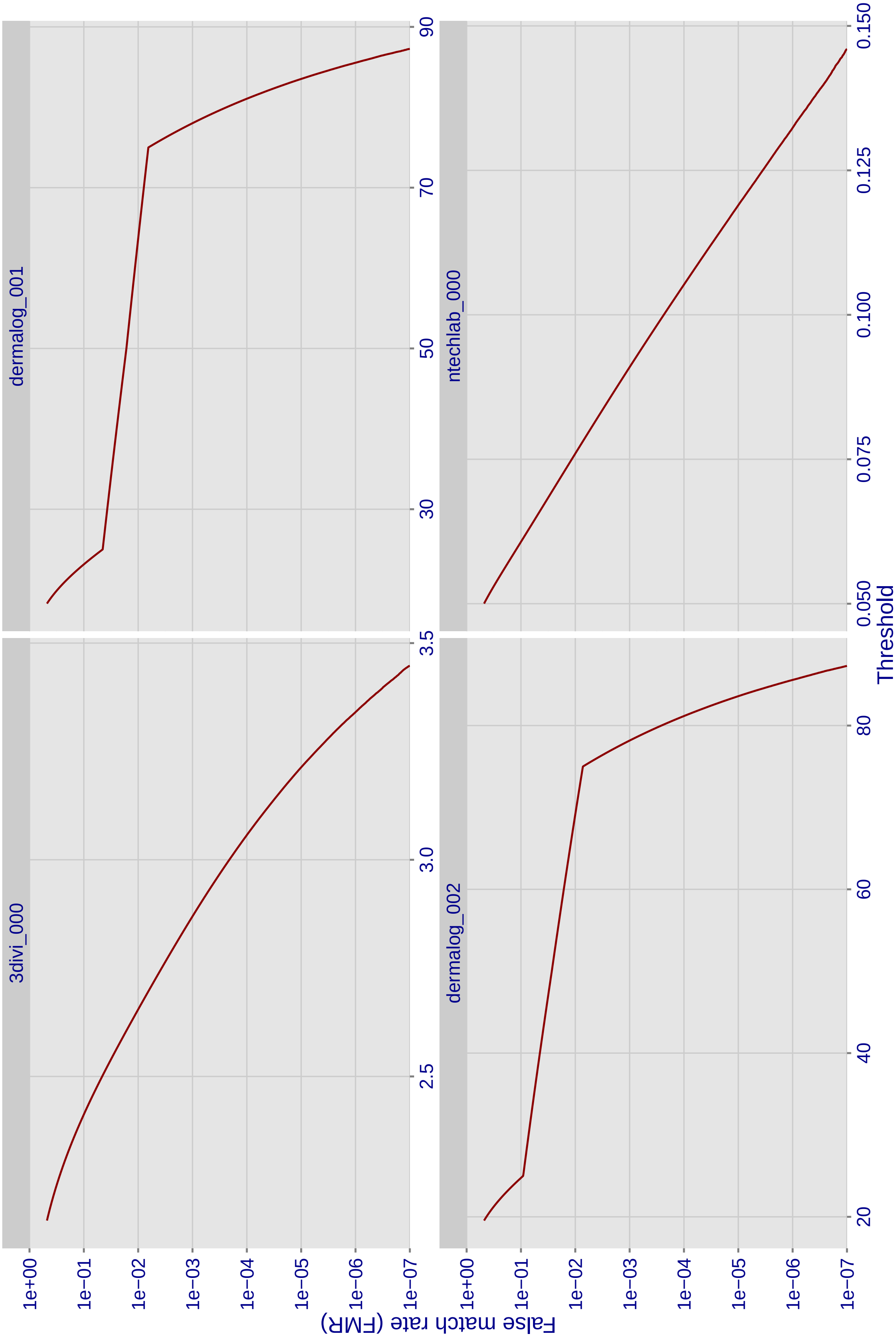

Figure 2: For the visa images, the false match calibration curves show false match rate vs. threshold.

3.2 Genuine distribution stability

3.2.1 Effect of birth place on the genuine distribution

Background: Faces change appearance throughout life. Face recognition algorithms have previously beeen reported to give better accuracy on older individuals (See NIST IR 8009).

Goal: To quantify false non-match rates (FNMR) as a function of age. We do not aim to quantify ageing effects here as the separation between two samples is limited to just a few years.

Methods: Thresholds are determined that give FMR = 0.001 and 0.0001 over the entire impostor set. Then FNMR is measured over 1000 bootstrap replications of the genuine scores. Only those countries with at least 140 individuals are included in the analysis.

Results: Figure 3 shows FNMR by country of birth for the two thresholds.

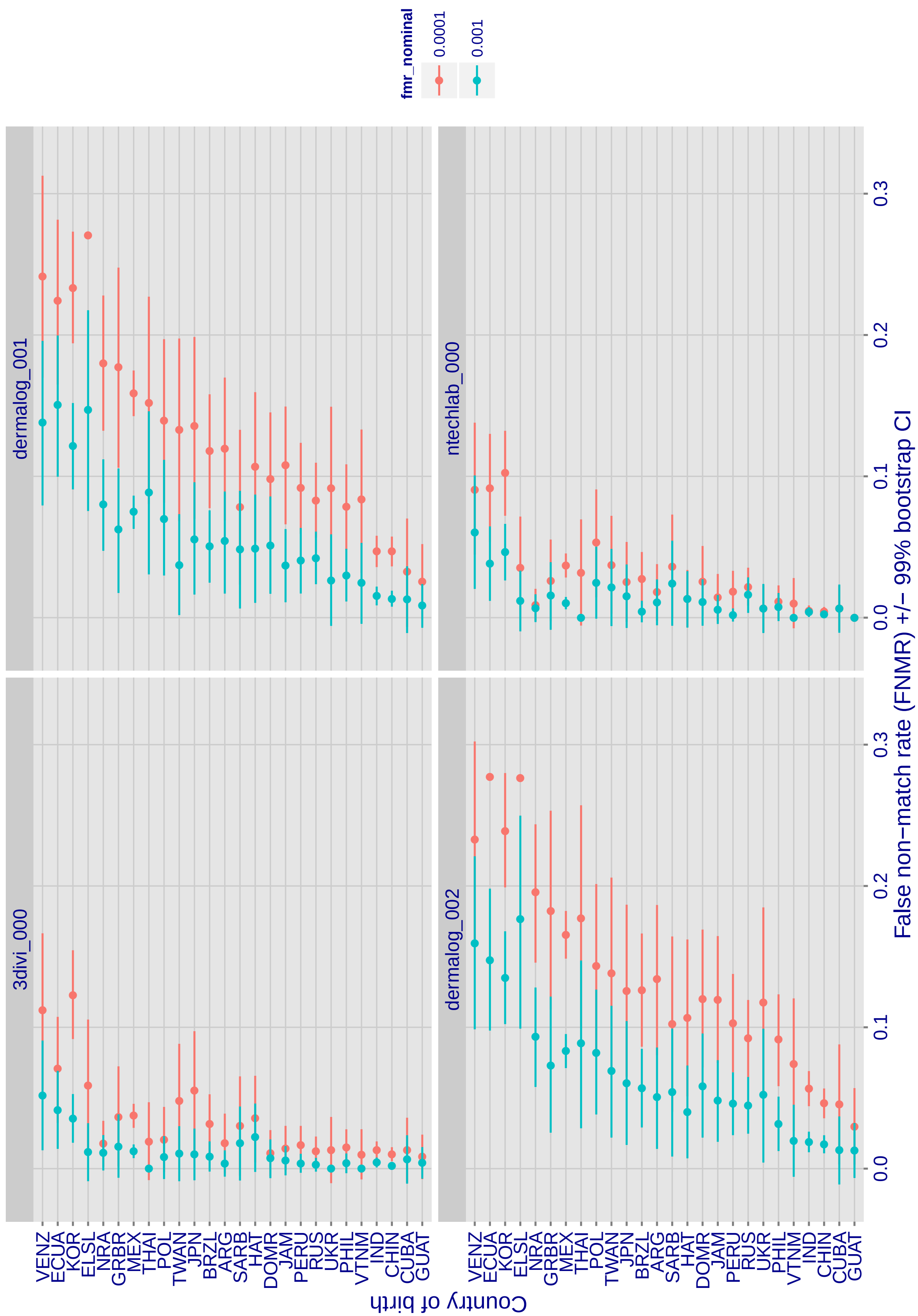

Figure 3: The figures shows an order of magnitude variation in FNMR across country of birth.

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

3.2.2 Effect of age on genuine subjects

Figure 4 shows how false non-match rates for genuine users, as a function of age group. The figure uses three thresholds corresponding to FMR(T) $= \{0.01, 0.001, 0.0001\}$.

The notable aspects are:

- ▷ Younger subjects give considerably higher FNMR. This is likely due rapid growth and change in facial appearance.
- ▷ FNMR trends down throughout life. The last bin, AGE > 72, is likely anomalous due to small sample size.

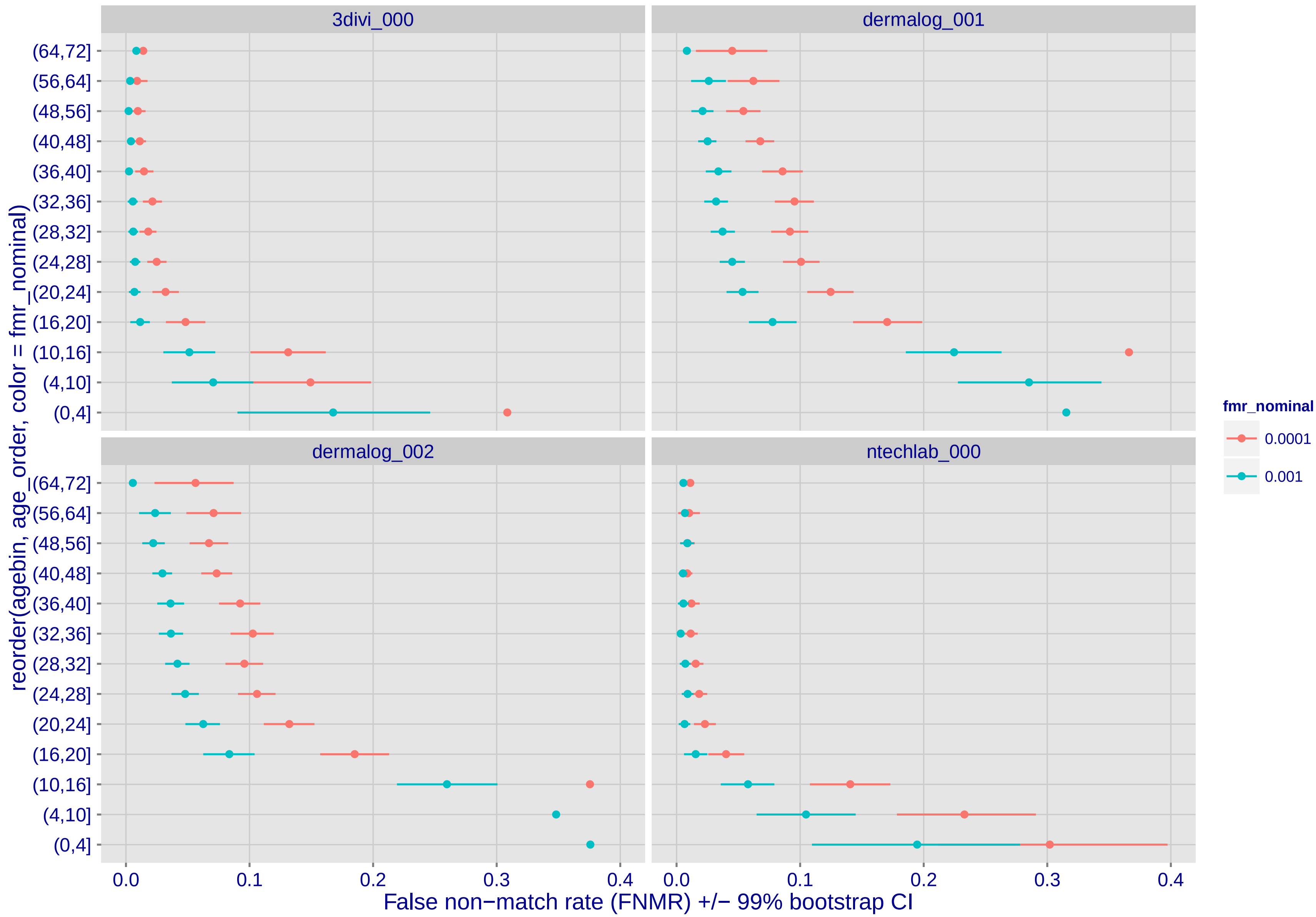

Figure 4: For the visa images, the dots show FNMR by age group for three operating thresholds corresponding to FMR = $\{0.01, 0.001, 0.0001\}$ *computed over all* $O(10^{10})$ *impostor scores. Given a pair of face images taken at different times, we assign a false non-match to the bin that is the arithmetic average of the subject's ages. This plot shows only the effect of age, not ageing. The number of comparisons in each bin is generally in the thousands. However the FNMR for the first and last bins are each computed over fewer than 150 comparisons.*

FNMR(T) "False non-match rate"
FMR(T) "False match rate"

3.3 Impostor distribution stability

3.3.1 Effect of birth place on the impostor distribution

Background: Facial appearance varies geographically, both in terms of skin tone, cranio-facial structure and size. This section addresses whether false match rates vary intra- and inter-regionally.

Goals:

- ▷ To show the effect of relative ages of the impostor and enrollee on false match rates.
- ▷ To determine whether some algorithms have better impostor distribution stability.

Methods:

- ▷ NIST defined 10 regions: Sub-Saharan Africa, South Asia, Polynesia, North Africa, Middle East, Europe, East Asia, Central and South America, Central Asia, and the Caribbean.
- ▷ NIST mapped each country of birth to a region. There is some arbitrariness to this. For example, Egypt could reasonably be assigned to the Middle East instead of North Africa. An alternative methodology could, for example, assign the Philippines to *both* Polynesia and East Asia.
- ▷ FMR is computed for cases where all face images of impostors born in region r_2 are compared with enrolled face images of persons born in region r_1.

$$\text{FMR}(r_1, r_2, T) = \frac{\sum_{i=1}^{N(r_1,r_2)} H(s_i - T)}{N_{r_1,r_2}} \tag{1}$$

where the same threshold, T, is used in all cells, and H is the unit step function. The threshold is set to give FMR(T) = 0.001 over the entire set of visa image impostor comparisons.

- ▷ This analysis is then repeated by country-pair, but only for those country pairs where both have at least 1000 images available. The countries[1] appear in the axes of graphs that follow.
- ▷ The mean number of impostor scores in any cross-region bin is 33 million. The smallest number of impostor scores in any bin is 135000, for Central Asia - North Africa. While these counts are large enough to support reasonable significance, the number of individual faces is much smaller, O($N^{0.5}$).
- ▷ The numbers of impostor scores in any cross-country bin is shown in Figure 12.

Results: Subsequent figures show heatmaps that use color to represent the base-10 logarithm of the false match rate. Red colors indicate high (bad) false match rates. Dark colors indicate benign false match rates. There are two series of graphs corresponding to aggregated geographical regions, and to countries. The notable observations are:

Caveats:

- ▷ The effects of variable impostor rates on one-to-many identification systems may well differ from what's implied by these one-to-one verification results. Two reasons for this are a) the enrollment galleries are usually imbalanced across countries of birth, age and sex; b) one-to-many identification algorithms often implement techniques aimed at stabilizing the impostor distribution. Further research is necessary.
- ▷ In principle, the effects seen in this subsection could be due to differences in the image capture process. We consider this unlikely since the effects are maintained across geography - e.g. Caribbean vs. Africa, or Japan vs. China.

[1] These are Argentina, Australia, Brazil, Chile, China, Costa Rica, Cuba, Czech Republic, Dominican Republic, Ecuador, Egypt, El Salvador, Germany, Ghana, Great Britain, Greece, Guatamala, Haiti, Hong Kong, Honduras, Indonesia, India, Israel, Jamaica, Japan, Kenya, Korea, Lebanon, Mexico, Malaysia, Nepal, Nigeria, Peru, Philippines, Pakistan, Poland, Romania, Russia, South Africa, Saudi Arabia, Thailand, Trinidad, Turkey, Taiwan, Ukraine, Venezuela, and Vietnam.

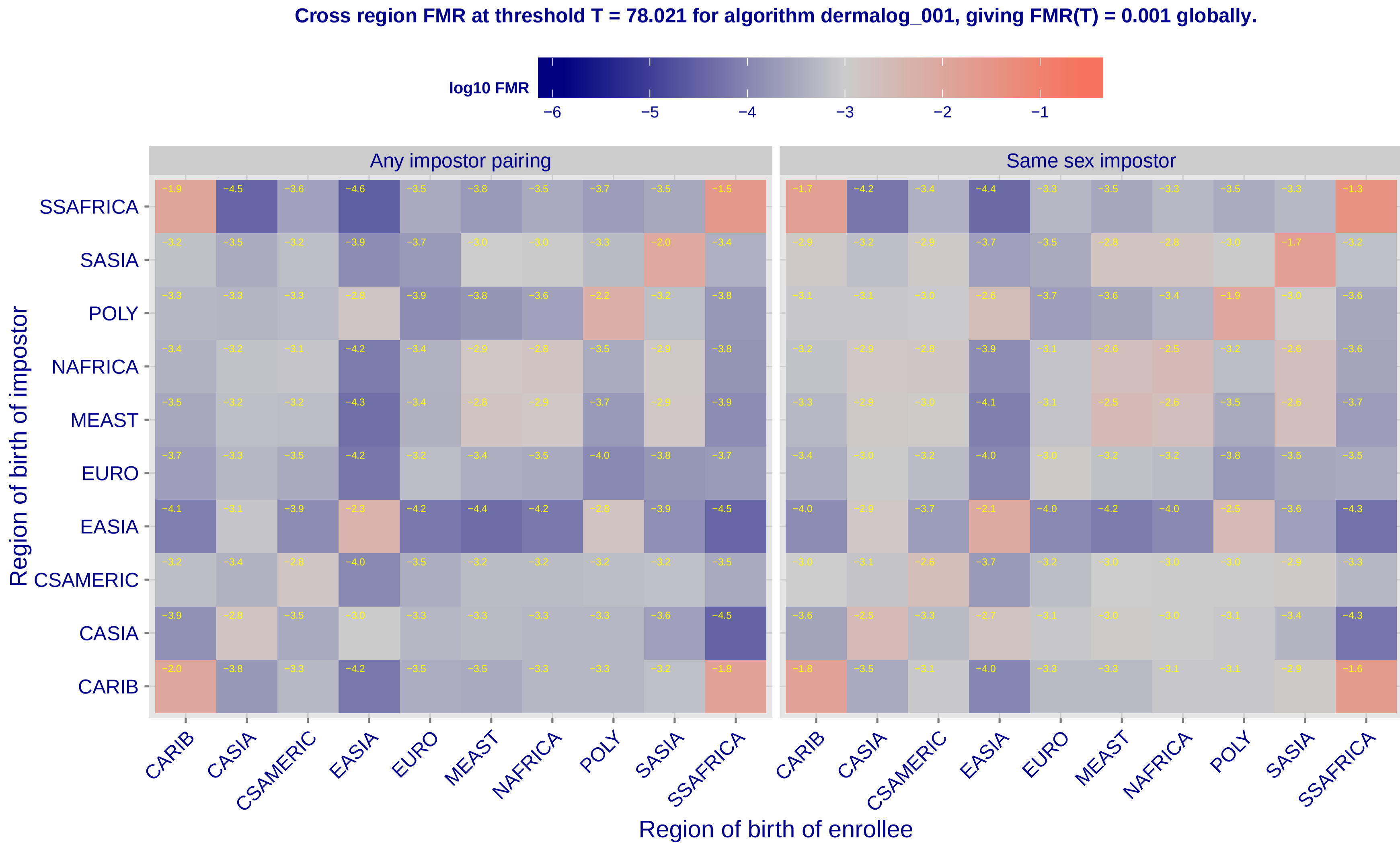

Figure 5: For algorithm dermalog-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who were born in the given region pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Light red colors present a security vulnerability to, for example, a passport gate.

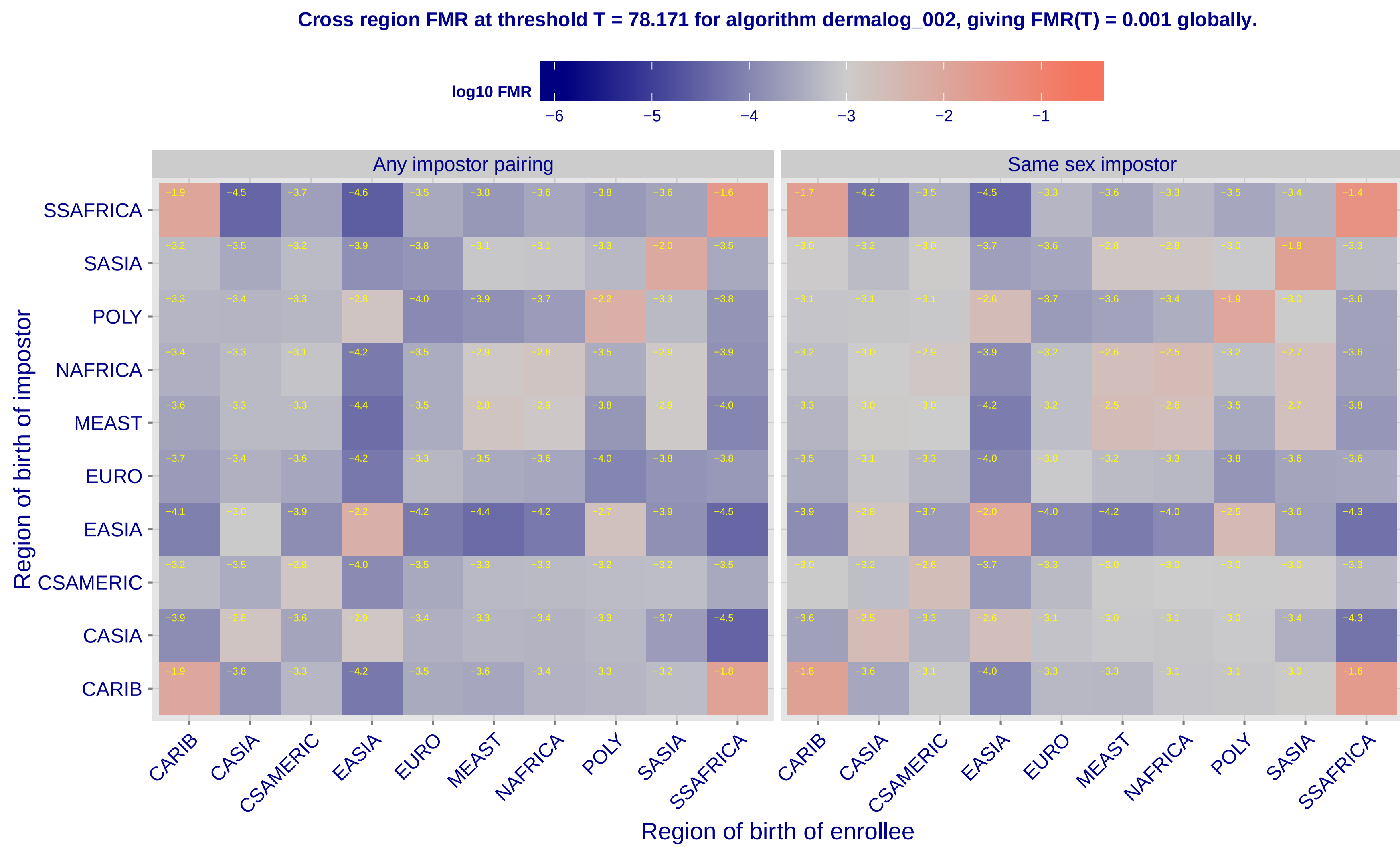

Figure 6: For algorithm dermalog-002 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who were born in the given region pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Light red colors present a security vulnerability to, for example, a passport gate.

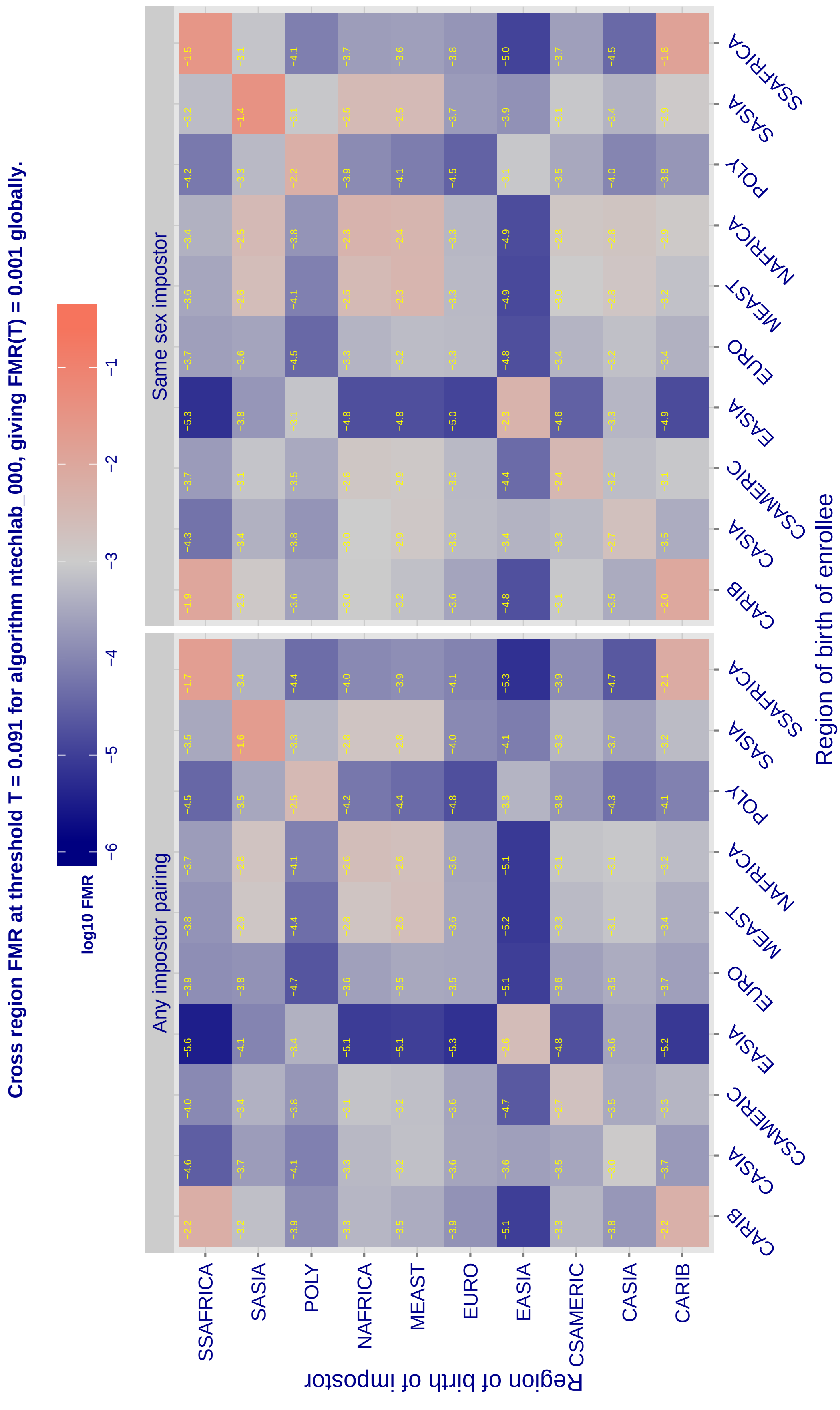

Figure 7: For algorithm ntechlab-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who were born in the given region pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Light red colors present a security vulnerability to, for example, a passport gate.

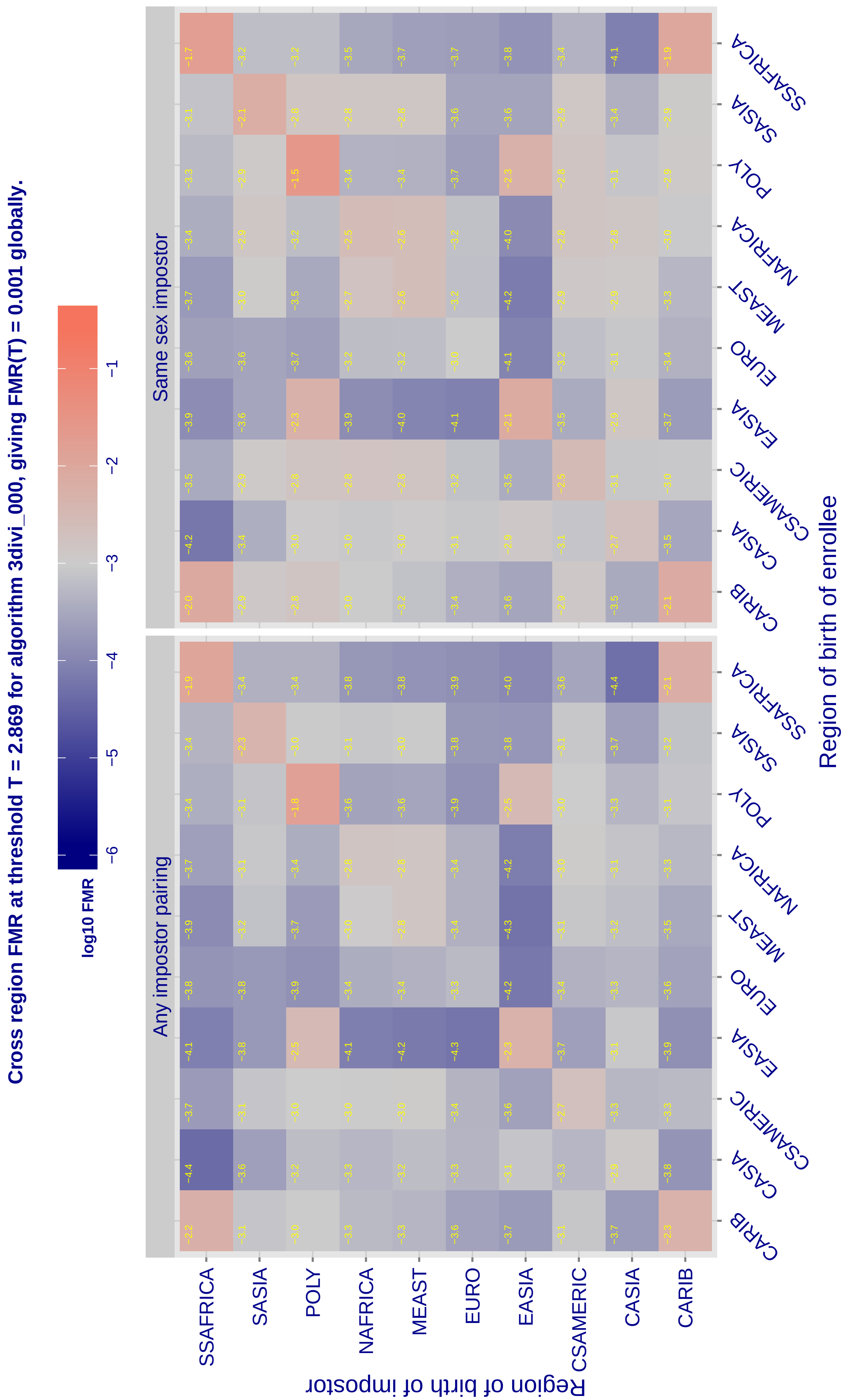

Figure 8: For algorithm 3divi-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who were born in the given region pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all O(10^{10}) impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Light red colors present a security vulnerability to, for example, a passport gate.

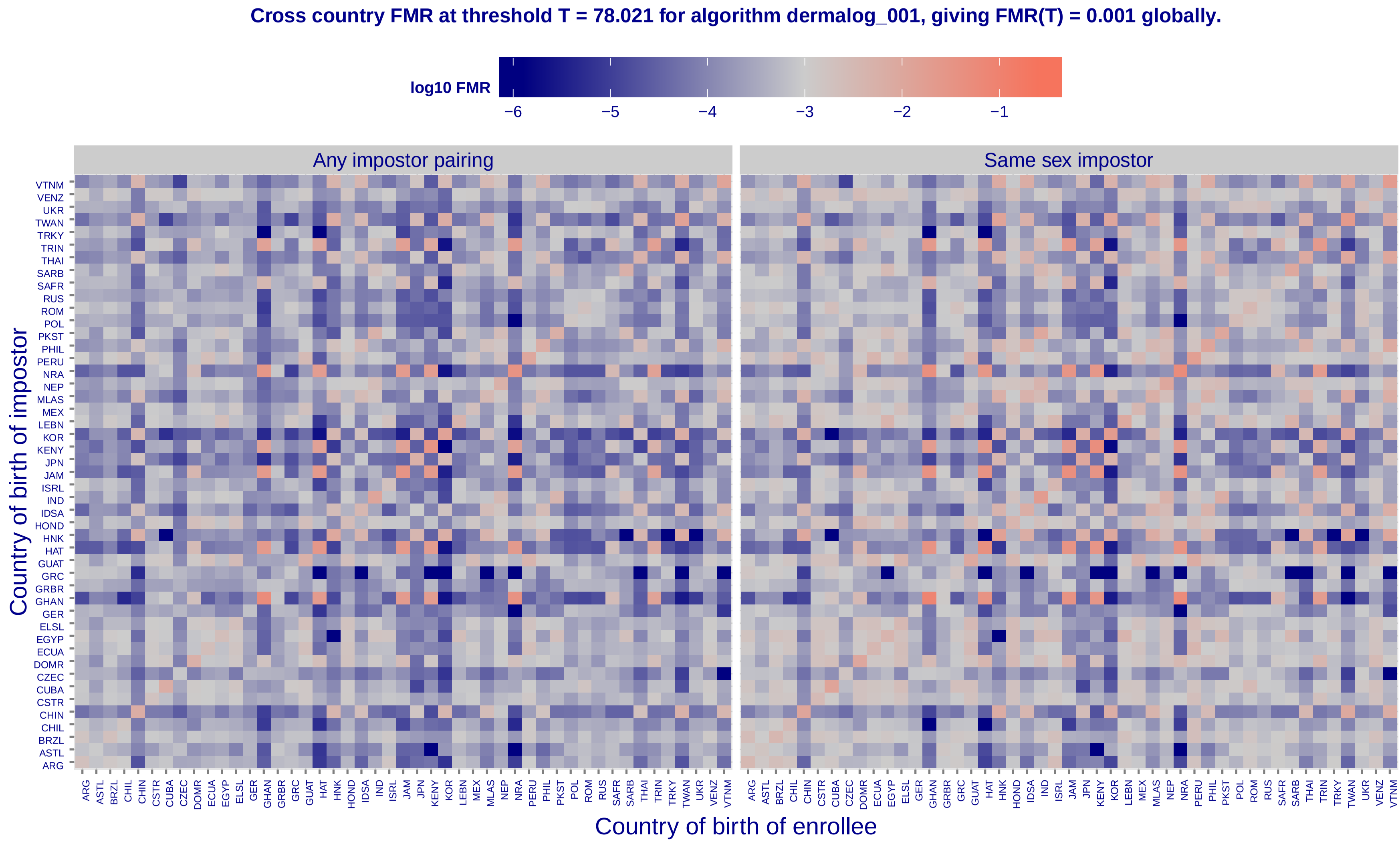

Figure 9: For algorithm dermalog-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Light red colors present a security vulnerability to, for example, a passport gate.

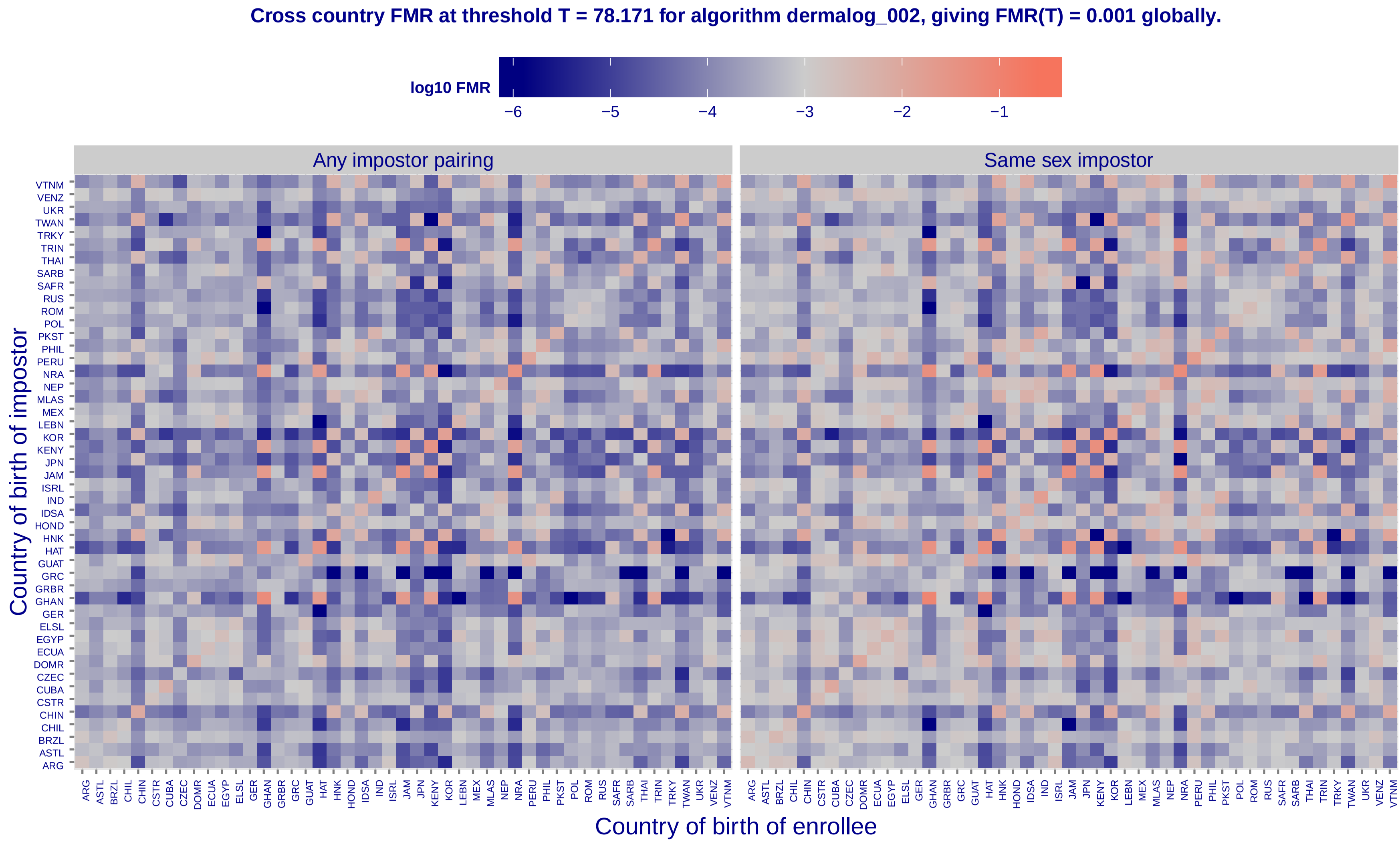

Figure 10: For algorithm dermalog-002 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Light red colors present a security vulnerability to, for example, a passport gate.

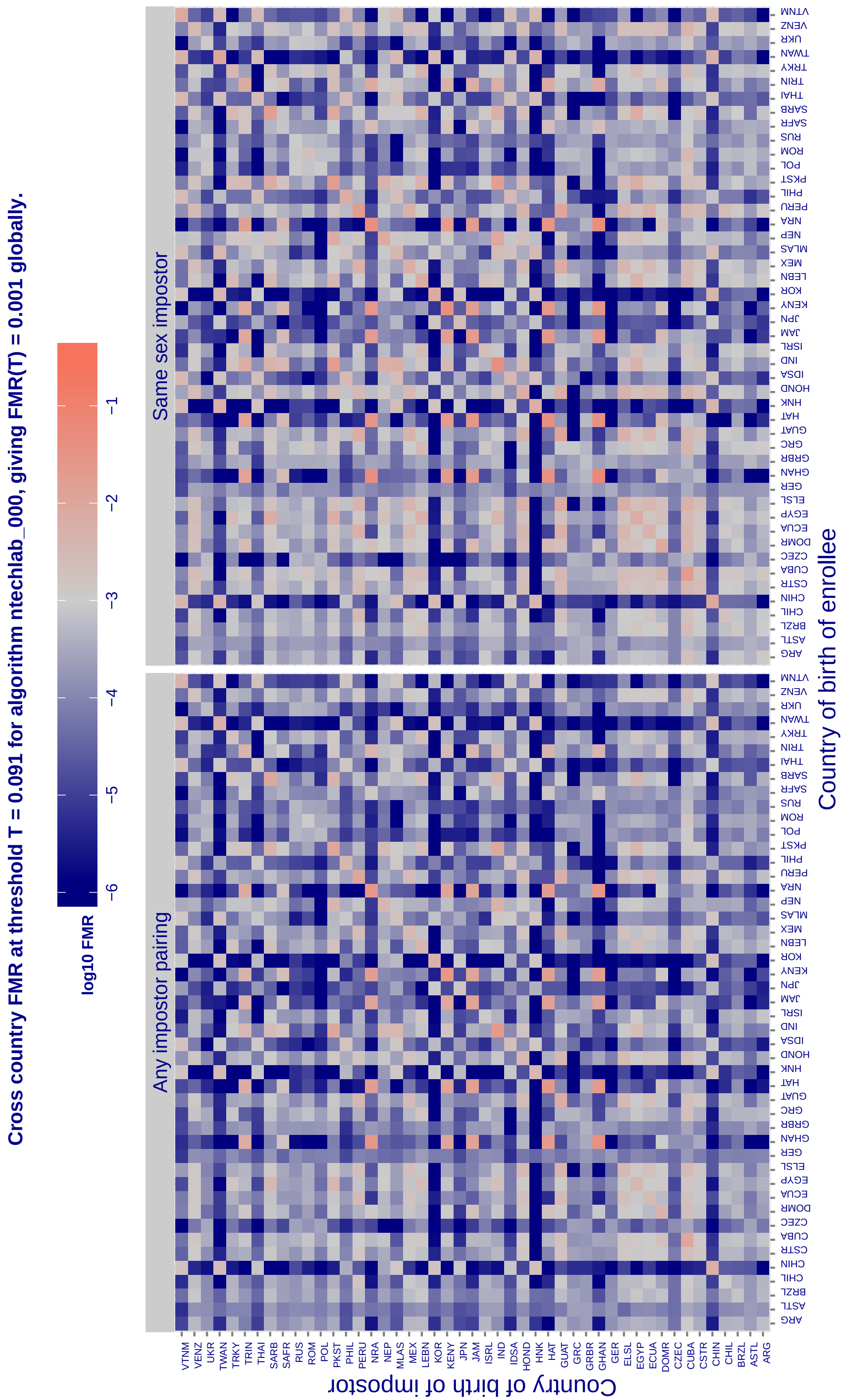

Figure 11: For algorithm ntechlab-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all O(10^{10}) impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Light red colors present a security vulnerability to, for example, a passport gate.

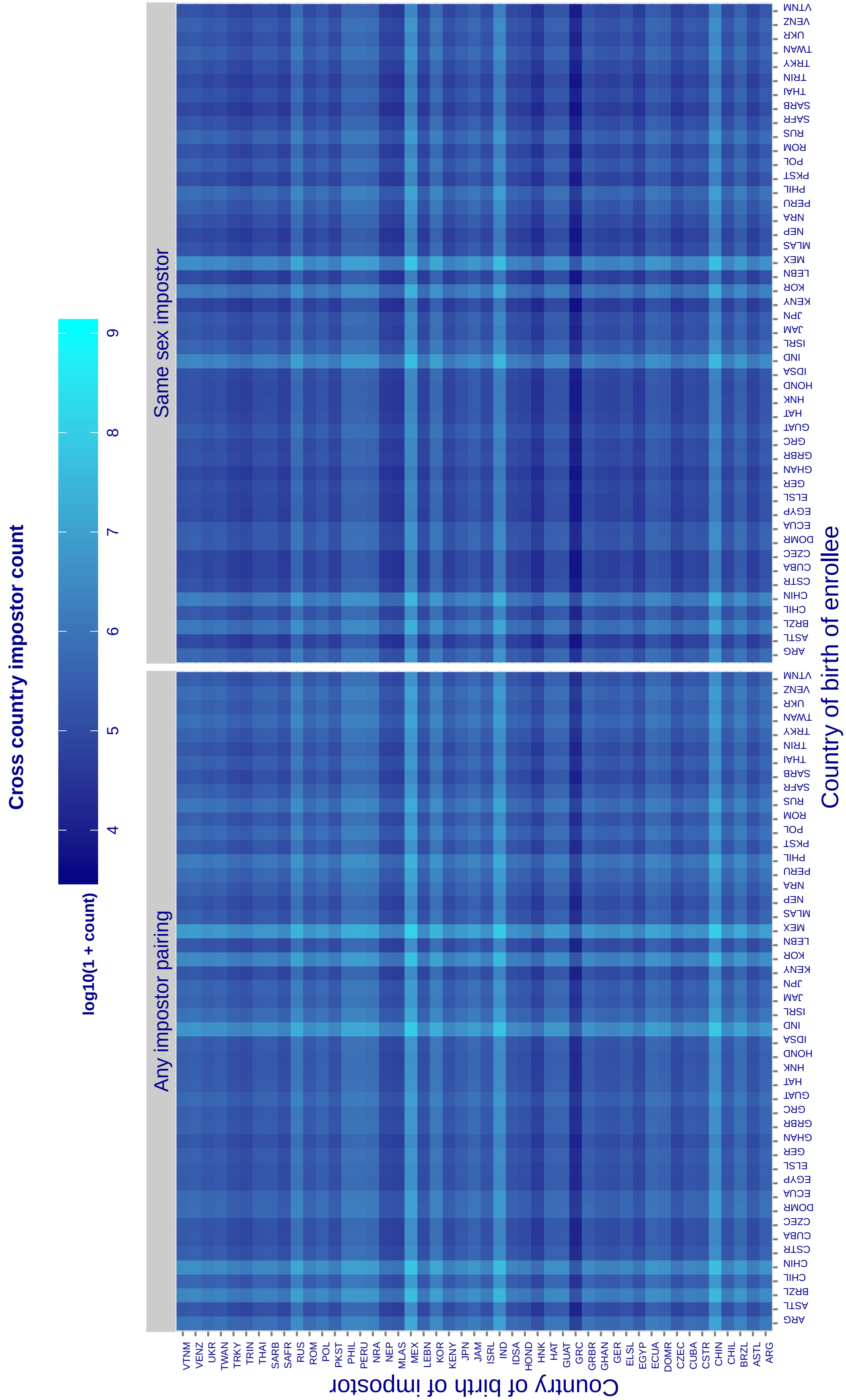

Figure 12: For visa images, the heatmap shows The count of impostor comparisons of faces from different individuals who were born in the given country pair.

3.3.2 Effect of age on impostors

Background: This section shows the effect of age on the impostor distribution. The ideal behaviour is that the age of the enrollee and the impostor would not affect impostor scores. This would support FMR stability over sub-populations.

Goals:

- ▷ To show the effect of relative ages of the impostor and enrollee on false match rates.
- ▷ To determine whether some algorithms have better impostor distribution stability.

Methods:

- ▷ Define 14 age group bins, spanning 0 to over 100 years old.
- ▷ Compute FMR over all impostor comparisons for which the subjects in the enrollee and impostor images have ages in two bins.
- ▷ Compute FMR over all impostor comparisons for which the subjects are additionally of the same sex, and born in the same geographic region.

Results:

The notable aspects are:

- ▷ Diagonal dominance: Impostors are more likely to be matched against their same age group.
- ▷ Same sex impostors are more successful. On the diagonal, an impostor is more likely to succeed by posing as someone of the same sex. If $\Delta \log_{10}$ FMR $= 0.2$, then same-sex FMR exceeds any-sex FMR by factor of $10^{0.2} = 1.6$. This is small relative to the age effect.
- ▷ Young children impostors give elevated FMR against young children. Older adult impostor give elevated FMR against older adults. These effects are quite large, for example if $\Delta \log_{10}$ FMR $= 1.0$ larger than a 32 year old, then these groups have higher FMR by a factor of $10^1 = 10$. This would imply an FMR above 0.01 for a nominal (global) FMR = 0.001.
- ▷ Algorithms vary.
- ▷ We computed the same quantities for a global FMR = 0.0001. The effects are similar.

Note the calculations in this section include impostors paired across all countries of birth.

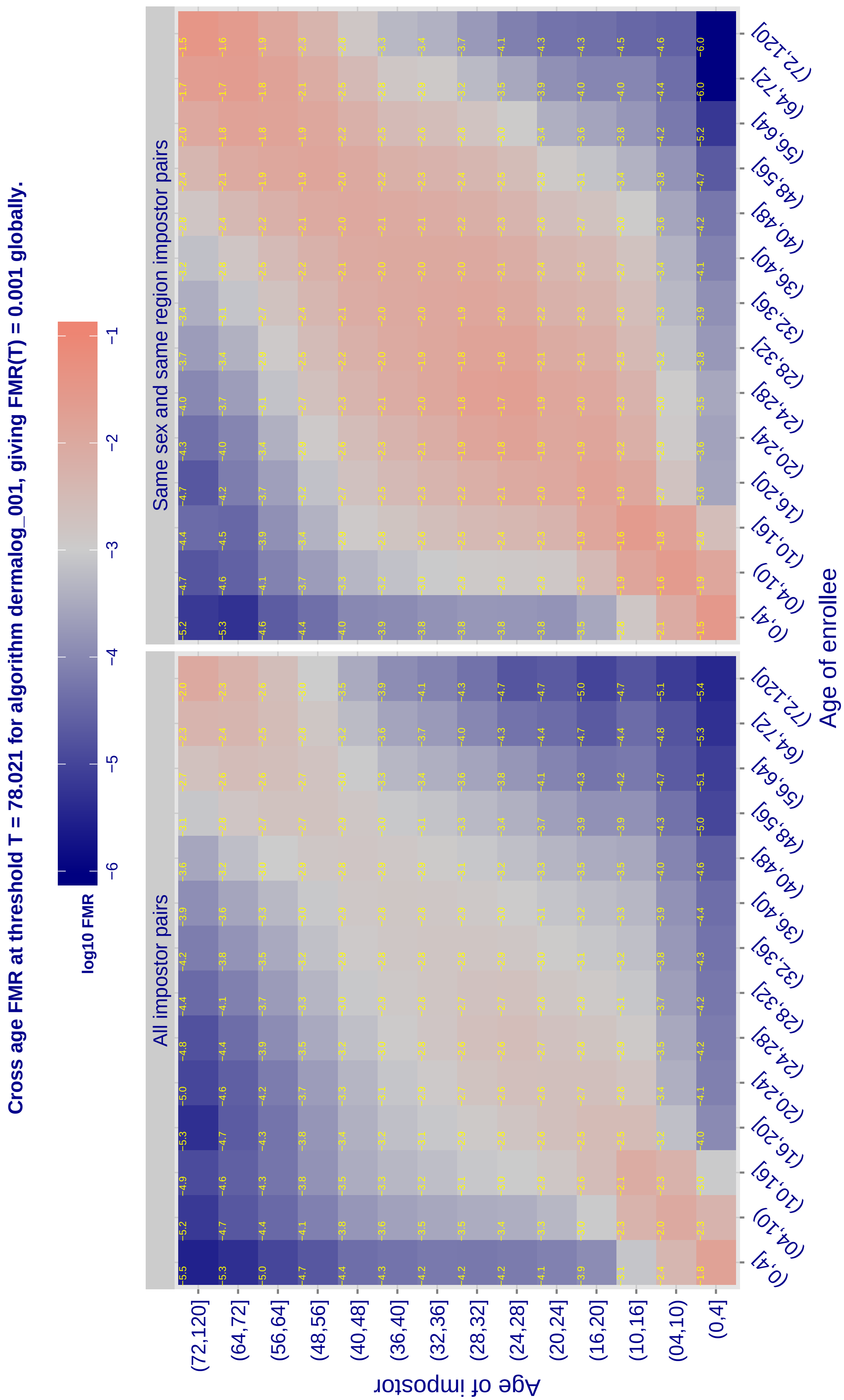

Figure 13: For algorithm dermalog-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all O(10^{10}) impostor comparisons. The text in each box it give the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

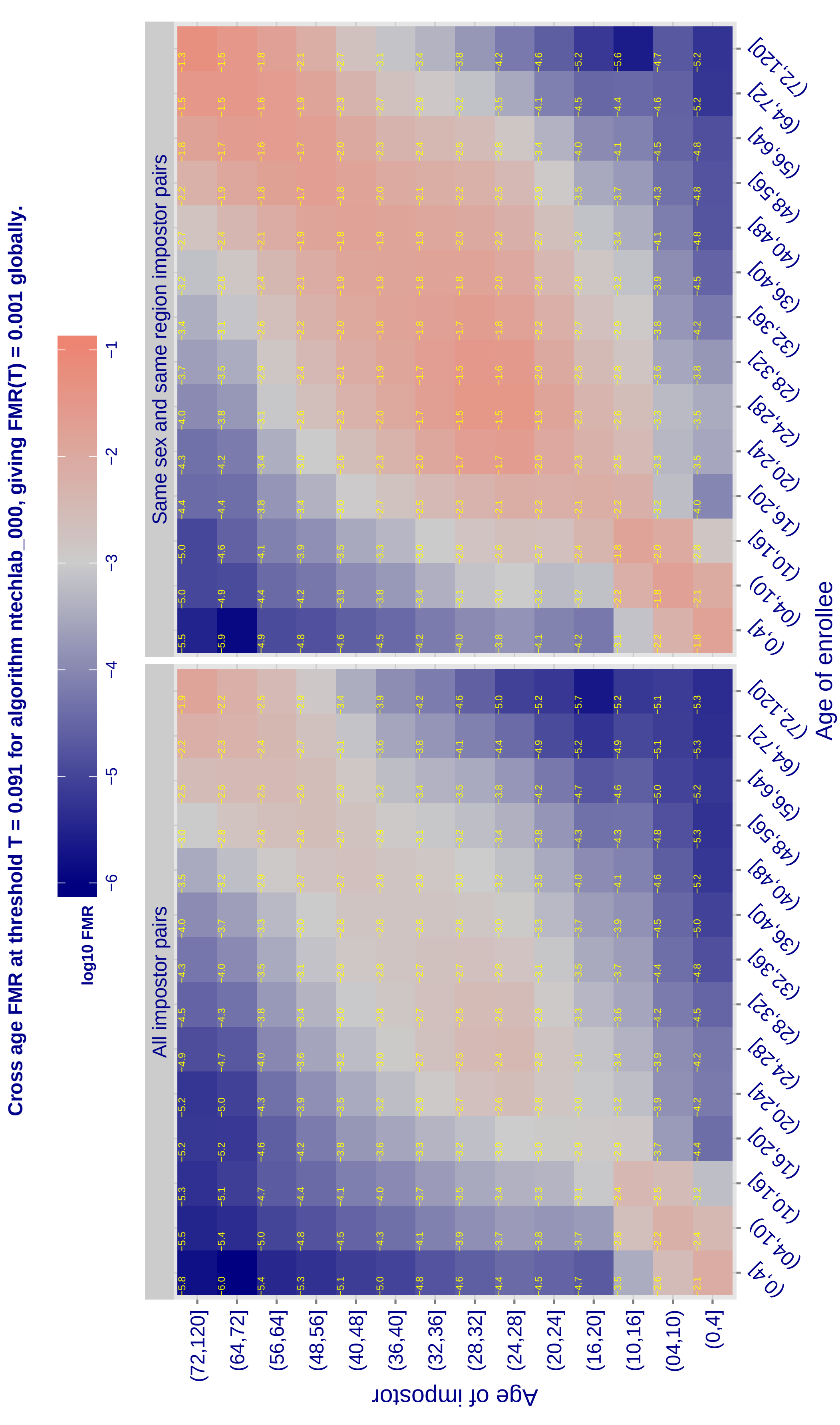

Figure 14: For algorithm ntechlab-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all O(10^{10}) impostor comparisons. The text in each box it give the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

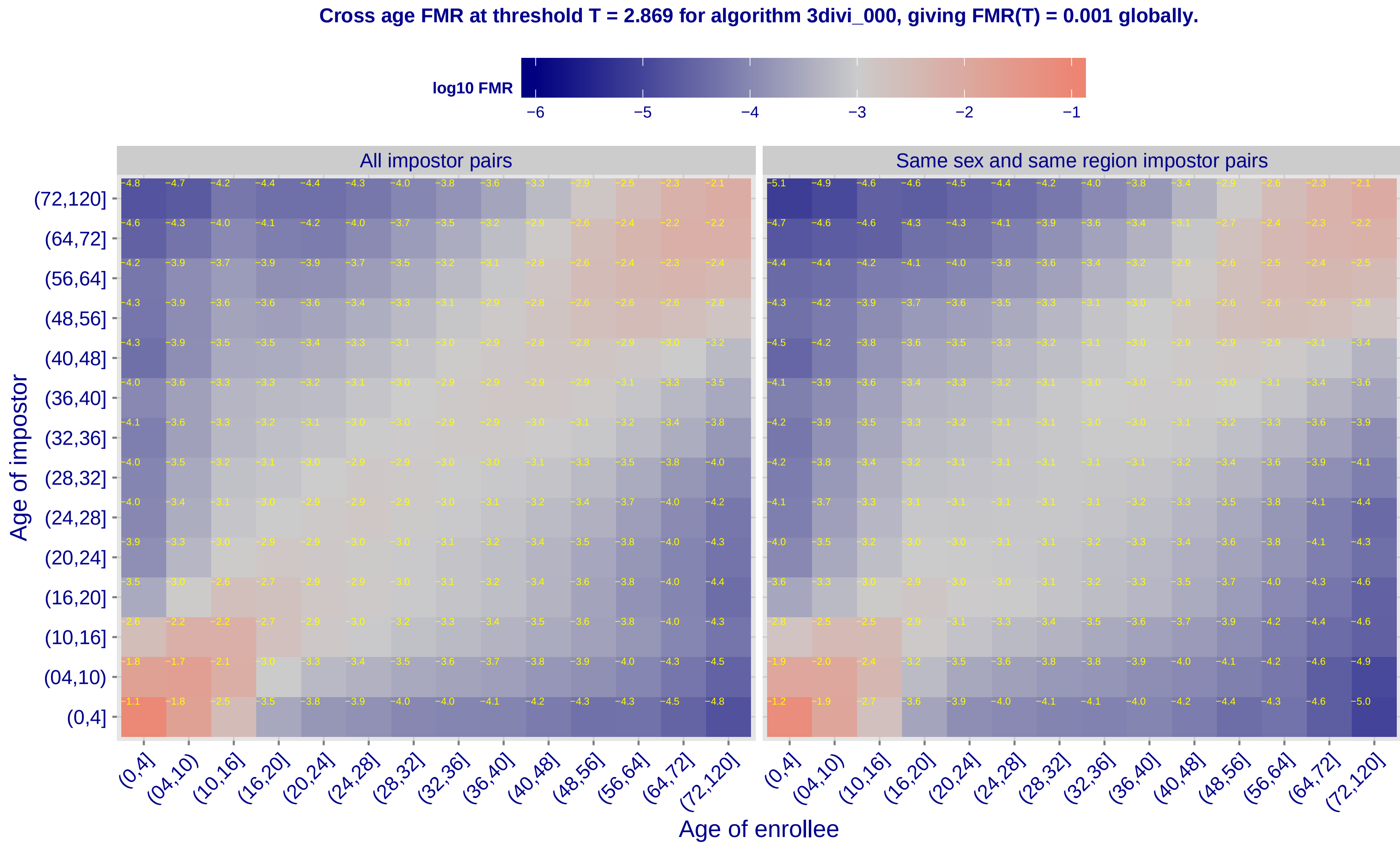

Figure 15: For algorithm 3divi-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give FMR = 0.001 over all O(10^{10}) impostor comparisons. The text in each box it give the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.