

NISTIR XXXX Draft

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

Patrick Grother
Mei Ngan
Kayee Hanaoka
*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>

DISCLAIMER

Specific hardware and software products identified in this report were used in order to perform the evaluations described in this document. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

ABOUT THIS REPORT

This report is a draft NIST Interagency Report, and is open for comment. It 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

DISCLAIMER	1
1 CHANGELOG	4
1.1 JUNE 19 2017	4
2 METRICS	7
2.1 CORE ACCURACY	7
3 DATASETS	8
3.1 CHILD EXPLOITATION IMAGES	8
3.2 VISA IMAGES	8
3.3 MUGSHOT IMAGES	8
3.4 SELFIE IMAGES	9
3.5 WEBCAM IMAGES	9
4 RESULTS	10
4.1 TEST GOALS	10
4.2 TEST DESIGN	10
4.3 FAILURE TO ENROL	12
4.4 RECOGNITION ACCURACY	12
4.5 GENUINE DISTRIBUTION STABILITY	26
4.5.1 EFFECT OF BIRTH PLACE ON THE GENUINE DISTRIBUTION	26
4.5.2 EFFECT OF AGE ON GENUINE SUBJECTS	28
4.6 IMPOSTOR DISTRIBUTION STABILITY	30
4.6.1 EFFECT OF BIRTH PLACE ON THE IMPOSTOR DISTRIBUTION	30
4.6.2 EFFECT OF AGE ON IMPOSTORS	67

List of Tables

1 ALGORITHM SUMMARY	5
2 FAILURE TO ENROL RATES	12

List of Figures

1 PERFORMANCE SUMMARY: FNMR VS. TEMPLATE TIME TRADEOFF	6
2 ERROR TRADEOFF CHARACTERISTIC: VISA IMAGES	13
3 ERROR TRADEOFF CHARACTERISTIC: MUGSHOT IMAGES	14
4 ERROR TRADEOFF CHARACTERISTIC: SELFIE IMAGES	15
5 ERROR TRADEOFF CHARACTERISTIC: SELFIE IMAGES	16
6 ERROR TRADEOFF CHARACTERISTIC: WILD IMAGES	17
7 ERROR TRADEOFF CHARACTERISTICS: CHILD EXPLOITATION IMAGES	18
8 ERROR TRADEOFF CHARACTERISTICS: CHILD EXPLOITATION IMAGES	19
9 SEX AND RACE EFFECTS: MUGSHOT IMAGES	20
10 SEX EFFECTS: VISA IMAGES	21
11 ERROR TRADEOFF CHARACTERISTIC: WILD IMAGES	22
12 FALSE MATCH RATE CALIBRATION: VISA IMAGES	23
13 FALSE MATCH RATE CONCENTRATION: VISA IMAGES	24
14 FALSE MATCH RATE CALIBRATION: MUGSHOT IMAGES	25
15 EFFECT OF COUNTRY OF BIRTH ON FNMR	27
16 EFFECT OF SUBJECT AGE ON FNMR	29
17 ALGORITHM 3DIVI-000 CROSS REGION FMR	32

18	ALGORITHM DERMALOG-001 CROSS REGION FMR	33
19	ALGORITHM DERMALOG-002 CROSS REGION FMR	34
20	ALGORITHM DIGITALBARRIERS-000 CROSS REGION FMR	35
21	ALGORITHM ITMO-001 CROSS REGION FMR	36
22	ALGORITHM NEUROTECHNOLOGY-000 CROSS REGION FMR	37
23	ALGORITHM NTECHLAB-000 CROSS REGION FMR	38
24	ALGORITHM NTECHLAB-001 CROSS REGION FMR	39
25	ALGORITHM RANKONE-000 CROSS REGION FMR	40
26	ALGORITHM RANKONE-001 CROSS REGION FMR	41
27	ALGORITHM SAMTECH-000 CROSS REGION FMR	42
28	ALGORITHM TONGYITRANS-001 CROSS REGION FMR	43
29	ALGORITHM TUPEL-001 CROSS REGION FMR	44
30	ALGORITHM VCOG-001 CROSS REGION FMR	45
31	ALGORITHM VIGILANTSOLUTIONS-000 CROSS REGION FMR	46
32	ALGORITHM VISIONLABS-001 CROSS REGION FMR	47
33	ALGORITHM VOCORD-001 CROSS REGION FMR	48
34	ALGORITHM VOCORD-002 CROSS REGION FMR	49
35	ALGORITHM YITU-000 CROSS REGION FMR	50
36	ALGORITHM 3DIVI-000 CROSS COUNTRY FMR	51
37	ALGORITHM DERMALOG-001 CROSS COUNTRY FMR	52
38	ALGORITHM DERMALOG-002 CROSS COUNTRY FMR	53
39	ALGORITHM NEUROTECHNOLOGY-000 CROSS COUNTRY FMR	54
40	ALGORITHM NTECHLAB-000 CROSS COUNTRY FMR	55
41	ALGORITHM NTECHLAB-001 CROSS COUNTRY FMR	56
42	ALGORITHM RANKONE-000 CROSS COUNTRY FMR	57
43	ALGORITHM RANKONE-001 CROSS COUNTRY FMR	58
44	ALGORITHM SAMTECH-000 CROSS COUNTRY FMR	59
45	ALGORITHM TONGYITRANS-001 CROSS COUNTRY FMR	60
46	ALGORITHM TUPEL-001 CROSS COUNTRY FMR	61
47	ALGORITHM VCOG-001 CROSS COUNTRY FMR	62
48	ALGORITHM VIGILANTSOLUTIONS-000 CROSS COUNTRY FMR	63
49	ALGORITHM VOCORD-001 CROSS COUNTRY FMR	64
50	ALGORITHM YITU-000 CROSS COUNTRY FMR	65
51	IMPOSTOR COUNTS FOR CROSS COUNTRY FMR CALCULATIONS	66
52	ALGORITHM 3DIVI-000 CROSS AGE FMR	68
53	ALGORITHM DERMALOG-001 CROSS AGE FMR	69
54	ALGORITHM DERMALOG-002 CROSS AGE FMR	70
55	ALGORITHM DIGITALBARRIERS-000 CROSS AGE FMR	71
56	ALGORITHM ITMO-001 CROSS AGE FMR	72
57	ALGORITHM NEUROTECHNOLOGY-000 CROSS AGE FMR	73
58	ALGORITHM NTECHLAB-000 CROSS AGE FMR	74
59	ALGORITHM NTECHLAB-001 CROSS AGE FMR	75
60	ALGORITHM RANKONE-000 CROSS AGE FMR	76
61	ALGORITHM RANKONE-001 CROSS AGE FMR	77
62	ALGORITHM SAMTECH-000 CROSS AGE FMR	78
63	ALGORITHM TONGYITRANS-001 CROSS AGE FMR	79
64	ALGORITHM TUPEL-001 CROSS AGE FMR	80
65	ALGORITHM VCOG-001 CROSS AGE FMR	81
66	ALGORITHM VIGILANTSOLUTIONS-000 CROSS AGE FMR	82
67	ALGORITHM VISIONLABS-001 CROSS AGE FMR	83
68	ALGORITHM VOCORD-001 CROSS AGE FMR	84
69	ALGORITHM VOCORD-002 CROSS AGE FMR	85
70	ALGORITHM YITU-000 CROSS AGE FMR	86

1 Changelog

1.1 June 19 2017

- ▷ Added five new algorithms, three of which remain in-process
- ▷ Added results for a child-exploitation dataset
- ▷ Added results for a “wild” dataset of images similar to non-cooperative photojournalism images
- ▷ Added Table 2 a standalone tabulation of failure to enrol rates
- ▷ Added Fig. 1 showing tradeoff between FNMR, template size, template generation time, and match duration.
- ▷ Added Fig. 13 showing how FMR is concentrated in certain images.
- ▷ Restated cross-region false match rates at nominal FMR = 0.0001 instead of 0.001
- ▷ Improved DET legends.

Developer	Short	Seq.	Validation	Config ¹	Template		GPU	Comparison Time (ns) ³	
Name	Name	Num.	Date	Data (KB)	Size (B)	Time (ms) ²		Genuine	Impostor
3DiVi	3divi	000	2017-03-16	169360	⁴ 512 ± 0	¹⁰ 285 ± 52	Yes	¹ 378 ± 20	² 375 ± 19
Dermalog	dermalog	001	2017-02-22	0	⁵ 1043 ± 0	⁹ 106 ± 1	No	⁸ 22160 ± 166	¹¹ 22131 ± 131
Dermalog	dermalog	002	2017-02-22	0	⁶ 1043 ± 0	² 81 ± 0	No	⁹ 22169 ± 114	¹⁰ 22105 ± 146
Digital Barriers	barriers	000	2017-05-31	157794	⁹ 2056 ± 0	⁴ 104 ± 0	No	⁴ 13232 ± 166	⁷ 13226 ± 146
ITMO University	itmo	001	2017-06-12	751338	¹⁹ 37997 ± 0	¹⁶ 870 ± 4	No	¹⁰ 29119 ± 1420	¹² 27817 ± 340
Neurotechnology	neurotech	000	2017-03-22	62129	¹⁷ 7148 ± 0	¹³ 611 ± 48	No	¹⁷ 74288 ± 2194	¹⁷ 72879 ± 2640
N-Tech Lab	ntech	000	2017-03-13	191530	¹¹ 2906 ± 1	⁹ 278 ± 13	No	¹¹ 30787 ± 142	¹³ 30846 ± 77
N-Tech Lab	ntech	001	2017-05-10	691296	¹⁶ 6744 ± 1	¹² 587 ± 11	No	¹⁵ 67692 ± 833	¹⁶ 67486 ± 244
Rank One Computing	rankone	000	2017-03-21	0	¹ 144 ± 0	³ 82 ± 9	No	¹³ 39932 ± 468	⁵ 8722 ± 171
Rank One Computing	rankone	001	2017-04-12	0	³ 208 ± 0	¹ 56 ± 4	No	¹⁶ 72754 ± 658	¹ 345 ± 29
Samtech InfoNet Limited	samtech	000	2017-05-02	109774	⁸ 2056 ± 0	⁷ 262 ± 2	No	³ 4550 ± 26	⁴ 4541 ± 28
TongYi Transportation Technology	tongyi	001	2017-04-01	625339	¹⁰ 2058 ± 0	¹¹ 310 ± 20	No	⁶ 17769 ± 74	⁹ 17750 ± 63
Tupel	tupel	001	2017-05-05	6347	¹⁴ 4142 ± 0	⁸ 273 ± 0	No	¹⁴ 56353 ± 618	¹⁵ 59204 ± 563
VCognition	vcog	001	2017-03-28	86103	¹² 4126 ± 0	⁶ 108 ± 17	Yes	⁵ 16320 ± 197	⁸ 16426 ± 425
Vigilant Solutions	vigilant	000	2017-03-30	352218	¹⁸ 31540 ± 0	¹⁷ 884 ± 23	No	⁷ 18201 ± 94	⁶ 13030 ± 83
Vision Labs	visionlabs	001	2017-06-12	343661	² 204 ± 0	¹⁹ 943 ± 8	No	² 1395 ± 45	³ 1148 ± 53
Vocord	vocord	001	2017-04-21	616989	¹⁵ 6194 ± 0	¹⁸ 908 ± 16	No	¹⁹ 1094730 ± 64282	¹⁹ 1107193 ± 66523
Vocord	vocord	002	2017-06-07	918292	⁷ 1330 ± 0	¹⁵ 782 ± 36	Yes	¹⁸ 83063 ± 517	¹⁸ 83072 ± 714
Shanghai Yitu Technology	yitu	000	2017-05-23	2211068	¹³ 4130 ± 0	¹⁴ 672 ± 2	No	¹² 35352 ± 114	¹⁴ 37848 ± 1773

Notes	
1	The size of configuration data does not capture static data included in the libraries. We do not include the size of the libraries because some algorithms include common ancilliary libraries for image processing (e.g. openCV) or numerical computation (e.g. blas).
2	The median template creation times are measured on Intel Xeon CPU E5-2630 v4 @ 2.20GHz processors or, in the case of GPU-enabled implementations, NVidia Tesla K40.
3	The median comparison durations, in nanoseconds, are estimated using std::chrono::high_resolution_clock which on the machine in (2) counts clock ticks of duration 1ns. Precision is somewhat worse than that however. The ± value is the median absolute deviation times 1.48 for Normal consistency.

Table 1: Summary of algorithms and properties included in this report. The red superscripts give ranking for the quantity in that column.

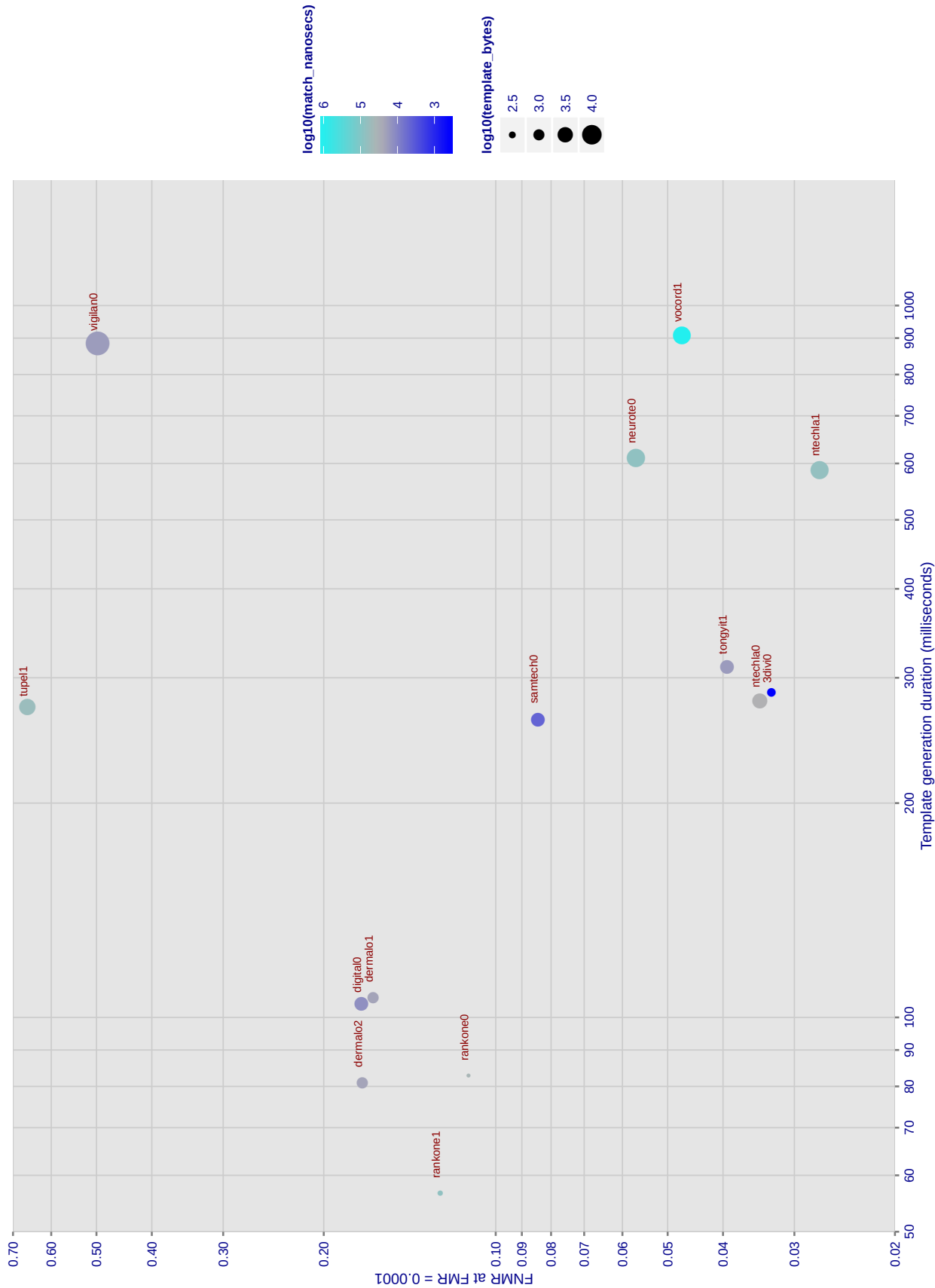


Figure 1: The points show false non-match rates (FNMR) versus the duration of the template generation operation. FNMR is the geometric mean of FNMR values for visa and mugshot images (from Figs. 2 and 3) at a false match rate (FMR) of 0.0001. Template generation time is a median estimated over 640 x 480 pixel portraits. The size of the points encodes template size - which span two orders of magnitude. The color of the points encodes one-to-one template comparison duration - which span three orders of magnitude.

2 Metrics

2.1 Core accuracy

Given a vector of N genuine scores, u , the false non-match rate (FNMR) is computed as the proportion below some threshold, T :

$$\text{FNMR}(T) = 1 - \frac{1}{N} \sum_i^N H(u_i - T) \quad (1)$$

where $H(x)$ is the unit step function, and $H(0)$ taken to be 1.

Similarly, given a vector of N impostor scores, v , the false match rate (FMR) is computed as the proportion above T :

$$\text{FMR}(T) = \frac{1}{N} \sum_i^N H(v_i - T) \quad (2)$$

The threshold, T , can take on any value. We typically generate a set of thresholds from quantiles of the observed impostor scores, v , as follows. Given some interesting false match rate range, $[\text{FMR}_L, \text{FMR}_U]$, we form a vector of K thresholds corresponding to FMR measurements evenly spaced on a logarithmic scale

$$T_k = Q_v(1 - \text{FMR}_k) \quad (3)$$

where Q is the quantile function, and FMR_k comes from

$$\log_{10} \text{FMR}_k = \log_{10} \text{FMR}_L + \frac{k}{K} [\log_{10} \text{FMR}_U - \log_{10} \text{FMR}_L] \quad (4)$$

Error tradeoff characteristics are plots of $\text{FNMR}(T)$ vs. $\text{FMR}(T)$. These are plotted with $\text{FMR}_U \rightarrow 1$ and FMR_L as low as is sustained by the number of impostor comparisons, N . This is somewhat higher than the “rule of three” limit $3/N$ because samples are not independent, due to re-use of images.

3 Datasets

3.1 Child exploitation images

- ▷ The number of images is $O(10^4)$.
- ▷ The number of subjects is $O(10^3)$.
- ▷ The number of subjects with two images $O(10^3)$.
- ▷ The images are operational. They are taken from ongoing investigations of child exploitation crimes. The images are arbitrarily unconstrained. Pose varies considerably around all three axes, including subject lying down. Resolution varies very widely. Faces can be occluded by other objects, including hair and hands. Lighting varies, although the images are intended for human viewing. Mis-focus is rare.
- ▷ The images are usually large from contemporary cameras. The mean interocular distance (IOD) is 70 pixels.
- ▷ The images are of subjects from several countries, with significant imbalance due to visa issuance patterns.
- ▷ The images are of subjects of children, from infancy to late adolescence.
- ▷ All of the images are live capture, none are scanned. Many have been cropped.
- ▷ When these images are input to the algorithm, they are labelled as being of type "EXPLOITATION" - see Table 4 of the FRVT API.

3.2 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. Pose is generally excellent.
- ▷ The images are of size 252x300 pixels. The mean interocular distance (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.
- ▷ Many of the images are live capture. A substantial number of the images are photographs of paper photographs.
- ▷ When these images are input to the algorithm, they are labelled as being of type "ISO" - see Table 4 of the FRVT API.

3.3 Mugshot images

- ▷ The number of images is $O(10^6)$.
- ▷ The number of subjects is $O(10^5)$.
- ▷ The number of subjects with two images $O(10^5)$.
- ▷ The images have geometry in reasonable conformance with the ISO/IEC 19794-5 Full Frontal image type.

- ▷ The images are of variable sizes. The median IOD is 104 pixels. The mean IOD is 123 pixels.
- ▷ The images are of subjects from the United States.
- ▷ The images are of adults.
- ▷ The images are all live capture.
- ▷ When these images are input to the algorithm, they are labelled as being of type "mugshot" - see Table 4 of the FRVT API.

3.4 Selfie images

- ▷ The number of images is below 500.
- ▷ The number of subjects is below 500.
- ▷ All subjects have a selfie image, and a portrait image.
- ▷ The portrait images are in reasonable conformance with the ISO/IEC 19794-5 Full Frontal image type.
- ▷ The selfie images vary: taken with camera above and below eye level, with one hand or two hands. Pitch angles vary more than yaw angles, which are frontal. Some perspective distortion is evident.
- ▷ The images have mean IOD of 140 pixels.
- ▷ The images are of subjects from the United States.
- ▷ The images are of adults.
- ▷ The images are all live capture.
- ▷ When these images are input to the algorithm, they are labelled as being of type "wild" - see Table 4 of the FRVT API.

3.5 Webcam images

- ▷ The number of images is below 1500.
- ▷ The number of subjects is below 1500.
- ▷ All subjects have a webcam image, and a portrait image.
- ▷ The portrait images are in reasonable conformance with the ISO/IEC 19794-5 Full Frontal image type.
- ▷ The webcam images are taken with camera at a typical head height, with mild pitch angles, low yaw angles, but some variation in range, such that low perspective distortion is sometimes evident.
- ▷ The images have mean IOD of 68 pixels (sd=12).
- ▷ The images are of subjects from the United States.
- ▷ The images are of adults.
- ▷ The images are all live capture.
- ▷ When these images are input to the algorithm, they are labelled as being of type "wild" - see Table 4 of the FRVT API.

4 Results

4.1 Test goals

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

4.2 Test design

Method: For visa images:

- ▷ 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. However, later analysis is conducted on subsets.
- ▷ The number of persons is $O(10^5)$.
- ▷ The number of images used to make 1 template is 1.
- ▷ The number of templates used to make each comparison score is two corresponding to simple one-to-one verification.

For mugshot images:

- ▷ The comparisons are of mugshot photos against mugshot photos.
- ▷ The number of genuine comparisons is $O(10^5)$.
- ▷ The number of impostor comparisons is $O(10^7)$.
- ▷ The comparisons are fully zero-effort, meaning impostors are paired without attention to sex, age or other covariates.
- ▷ The number of persons is $O(10^6)$.
- ▷ The number of images used to make 1 template is 1.
- ▷ The number of templates used to make each comparison score is two corresponding to simple one-to-one verification.

For selfie images:

- ▷ The comparisons are of selfie photos against portrait photos.
- ▷ The number of genuine comparisons is $O(10^2)$.
- ▷ The number of impostor comparisons is $O(10^8)$ selfies are compared with portraits of $O(10^6)$ other subjects.
- ▷ The comparisons are fully zero-effort, meaning impostors are paired without attention to sex, age or other covariates.
- ▷ The number of persons is $O(10^6)$.

- ▷ The number of images used to make 1 template is 1.
- ▷ The number of templates used to make each comparison score is two corresponding to simple one-to-one verification.

For webcam images:

- ▷ The comparisons are of webcam photos against portrait photos.
- ▷ The number of genuine comparisons is $O(10^3)$.
- ▷ The number of impostor comparisons is $O(10^9)$ webcams are compared with portraits of $O(10^6)$ other subjects.
- ▷ The comparisons are fully zero-effort, meaning impostors are paired without attention to sex, age or other covariates.
- ▷ The number of persons is $O(10^6)$.
- ▷ The number of images used to make 1 template is 1.
- ▷ The number of templates used to make each comparison score is two corresponding to simple one-to-one verification.

For child exploitation images:

- ▷ The comparisons are of unconstrained child exploitation photos against others of the same type.
- ▷ The number of genuine comparisons is $O(10^4)$.
- ▷ The number of impostor comparisons is $O(10^7)$.
- ▷ The comparisons are fully zero-effort, meaning impostors are paired without attention to sex, age or other covariates.
- ▷ The number of persons is $O(10^3)$.
- ▷ The number of images used to make 1 template is 1.
- ▷ The number of templates used to make each comparison score is two corresponding to simple one-to-one verification.
- ▷ We produce two performance statements. First, is a DET as used for visa and mugshot images. The second is a cumulative match characteristic (CMC) summarizing a simulated one-to-many search process. This is done as follows.
 - We regard M enrollment templates as items in a gallery.
 - These M templates come from $M > N$ individuals, because multiple images of a subject are present in the gallery under separate identifiers.
 - We regard the verification templates as search templates.
 - For each search we compute the rank of the highest scoring mate.
 - This process should properly be conducted with a 1:N algorithm, such as those tested in NIST IR 8009. We use the 1:1 algorithms in a simulated 1:N mode here to a) better reflect what a child exploitation analyst does, and b) to do show algorithm efficacy is better than that revealed in the verification DETs.

Algorithm Name	Failure to Enrol Rate ¹									
	CHILD-EXPLOIT		MUGSHOT		SELFIES		VISA		WEBCAM	
3divi-000	0.2019	9	0.0019	11	0.0202	10	0.0008	7	0.0020	9
dermalog-001	0.9109	16	0.0045	14	0.0954	16	0.0013	9	0.0471	15
dermalog-002	0.9109	17	0.0045	15	0.0954	17	0.0013	10	0.0471	16
digitalbarriers-000	0.5469	13	0.0043	13	0.0925	15	0.0019	16	0.0184	12
itmo-001	-	19	-	19	-	19	0.0047	21	-	19
neurotechnology-000	0.0000	2	0.0000	1	0.0000	3	0.0000	1	0.0000	4
ntechlab-000	0.2496	10	0.0015	9	0.0058	9	0.0016	13	0.0007	6
ntechlab-001	0.0926	7	0.0009	5	0.0029	5	0.0005	5	0.0007	5
rankone-000	0.0187	5	0.0005	3	0.0000	4	0.0003	4	0.0007	7
rankone-001	0.0012	4	0.0001	2	0.0000	2	0.0000	2	0.0000	3
samtech-000	0.5474	14	0.0052	16	0.0491	13	0.0042	20	0.0252	13
tongyitrans-001	0.0000	3	0.0068	17	0.0462	12	0.0040	19	0.0055	10
tupel-001	-	19	0.0030	12	0.0000	1	0.0018	15	0.0000	2
vcog-001	0.1579	8	-	19	0.0058	8	0.0018	14	0.0000	1
vigilantsolutions-000	0.5580	15	0.0018	10	0.0462	11	0.0007	6	0.0109	11
visionlabs-001	-	19	-	19	-	19	0.0014	12	-	19
vocord-001	0.4732	12	0.0158	18	0.0520	14	0.0038	18	0.0348	14
vocord-002	-	19	0.0015	7	0.0029	7	0.0037	17	-	19
yitu-000	0.3475	11	0.0015	8	0.0029	6	0.0013	11	0.0014	8

Table 2: FTE is the proportion of failed template generation attempts. Failures can occur because the software throws an exception, or because the software electively refuses to process the input image. This would typically occur if a face is not detected. FTE is measured as the number of function calls that give a non-zero error code, OR that give a “small” template. This is defined as one whose size is less than 0.3 times the median template size. This second rule is needed because some algorithms incorrectly fail to return a non-zero error code when template generation fails.

The effects of FTE are included in the accuracy results later in this report by regarding any template comparison that involves a failed template is taken to produce a low similarity score.

4.3 Failure to enrol

4.4 Recognition accuracy

Core algorithm accuracy is stated via:

- ▷ The visa image DETs of Figure 2;
- ▷ The mugshot DETs of Figure 3 ;
- ▷ The selfie-portrait DETs of Figure 4;
- ▷ The webcam-portrait DETs of Figure 5;
- ▷ The child-exploitation DET of Figure 7;
- ▷ The child-exploitation CMC of Figure 8.

Figure 12 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.

Figure 14 likewise shows FMR(T) but for mugshots, and specially four subsets of the population.

Note that in both the mugshot and visa sets false match rates vary with the ethnicity, age, and sex, of the enrollee and impostor - see section 4.6.

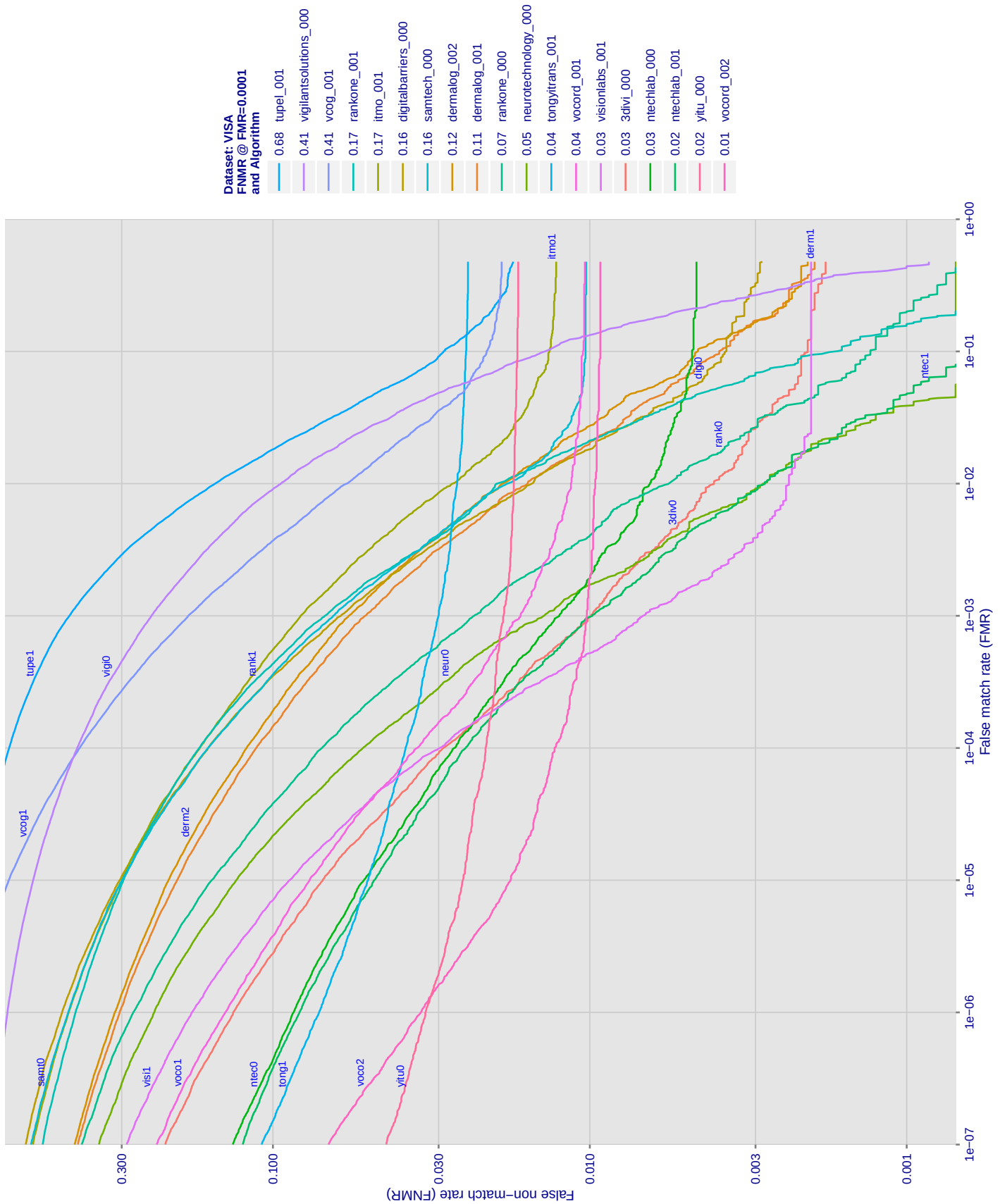


Figure 2: For the visa images, detection error tradeoff (DET) characteristics showing false non-match rate vs. false match rate plotted parametrically on threshold, T . The scales are logarithmic in order to show many decades of FMR.

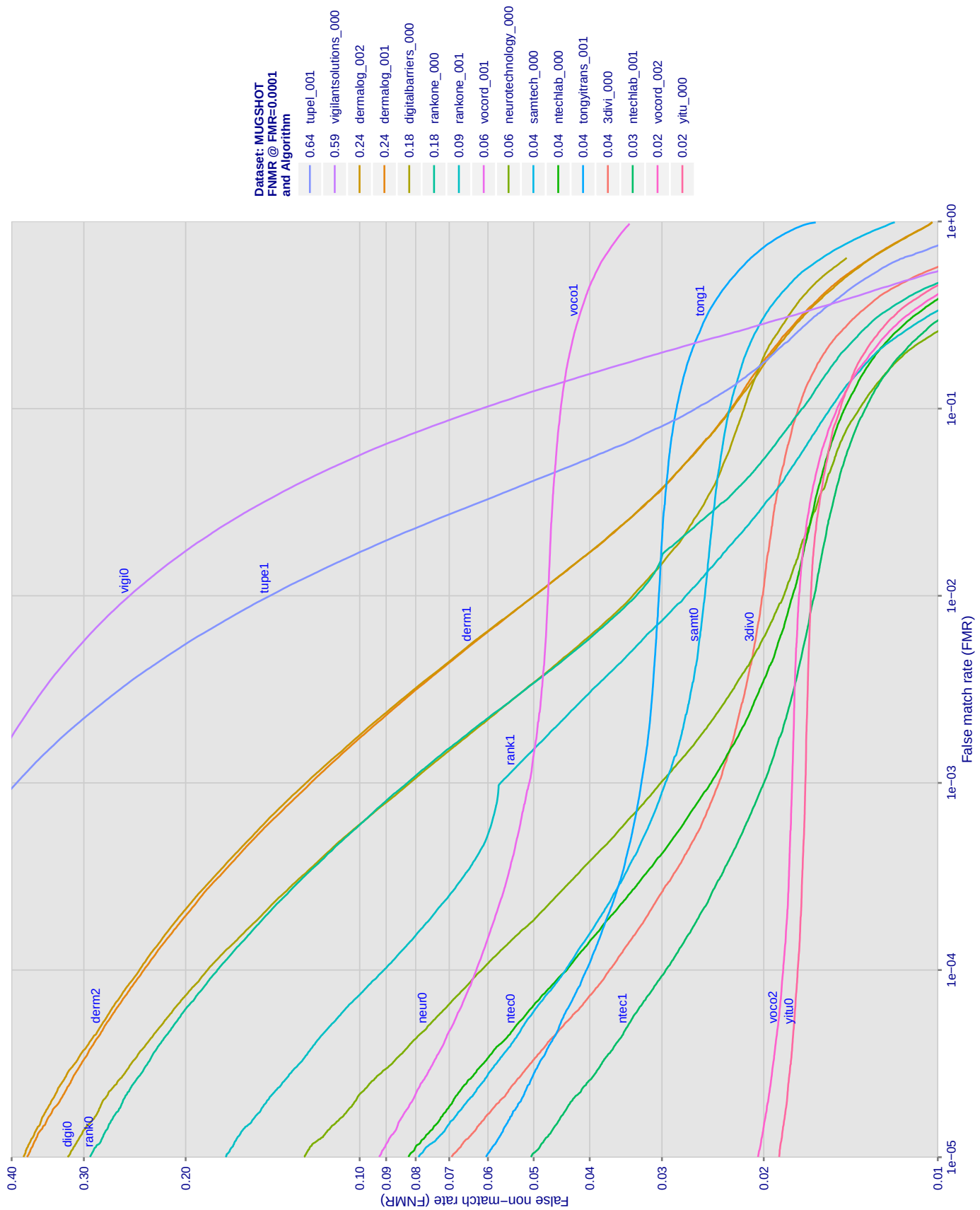


Figure 3: For the mugshot images, detection error tradeoff (DET) characteristics showing false non-match rate vs. false match rate plotted parametrically on threshold, T . The scales are logarithmic in order to show decades of FNMR.

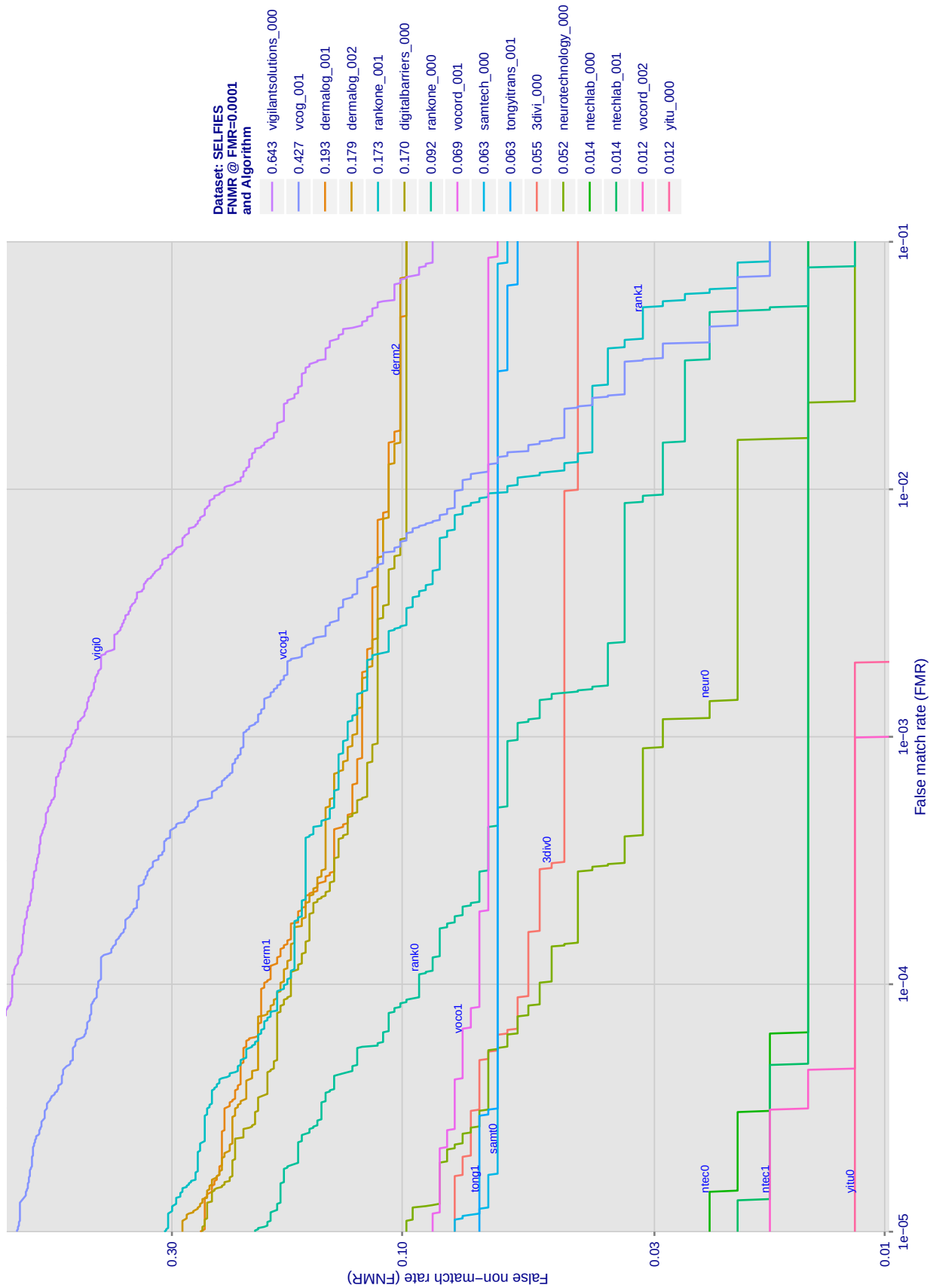


Figure 4: For the selfie-to-portrait comparisons, detection error tradeoff (DET) characteristics showing false non-match rate vs. false match rate plotted parametrically on threshold, T . The scales are logarithmic in order to show several decades of FMR. **Caution: The FNMR values here are optimistic statements of accuracy because the image pairs were collected on the same day. This is known across biometrics to give better accuracy, and is operationally relevant only in special cases.**

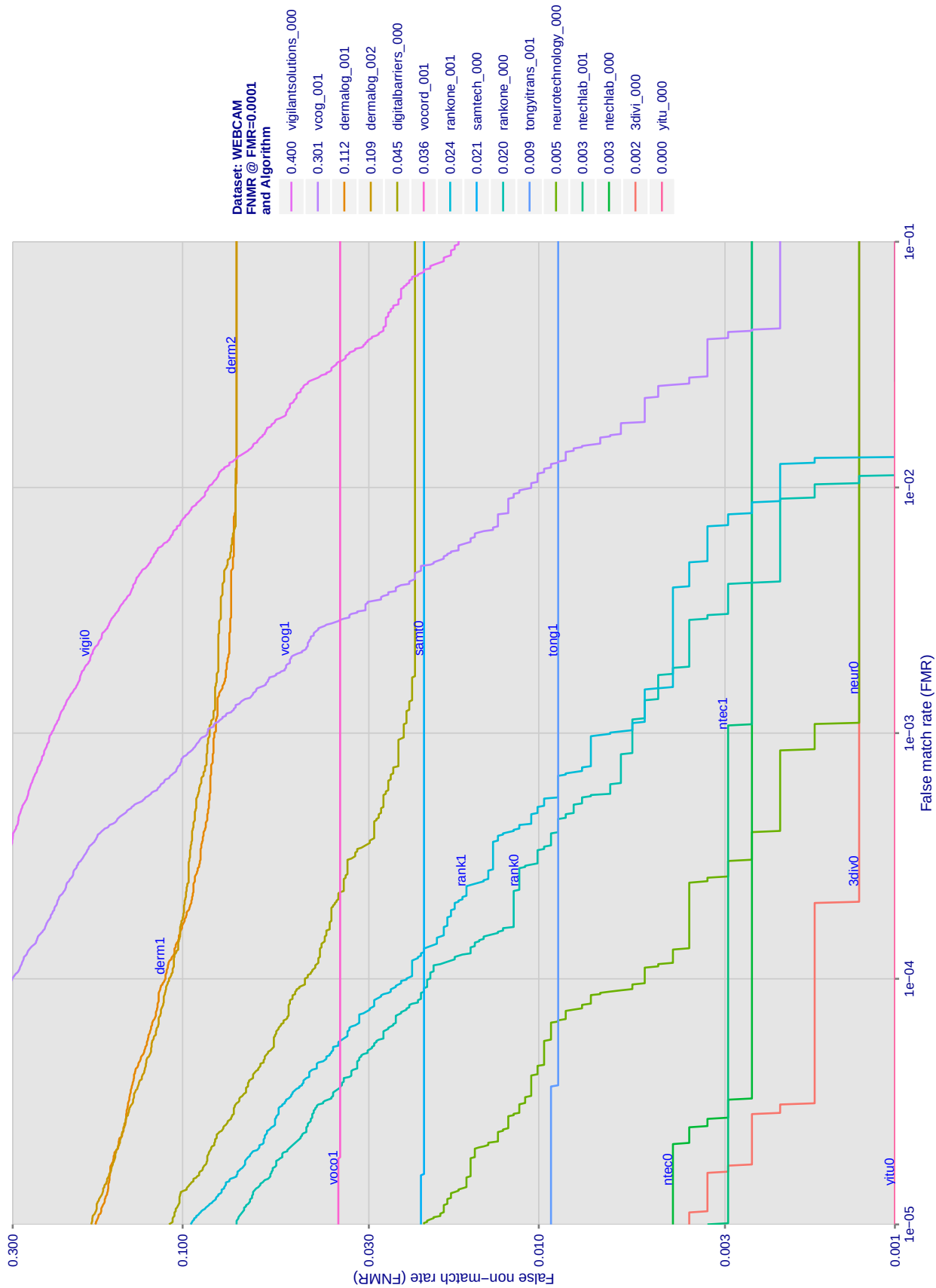


Figure 5: For the webcam-to-portrait comparisons, detection error tradeoff (DET) characteristics showing false non-match rate vs. false match rate plotted parametrically on threshold, T. The scales are logarithmic in order to show several decades of FMR. **Caution: The FNMR values here are optimistic statements of accuracy because the image pairs were collected on the same day. This is known across biometrics to give better accuracy, and is operationally relevant only in special cases.**

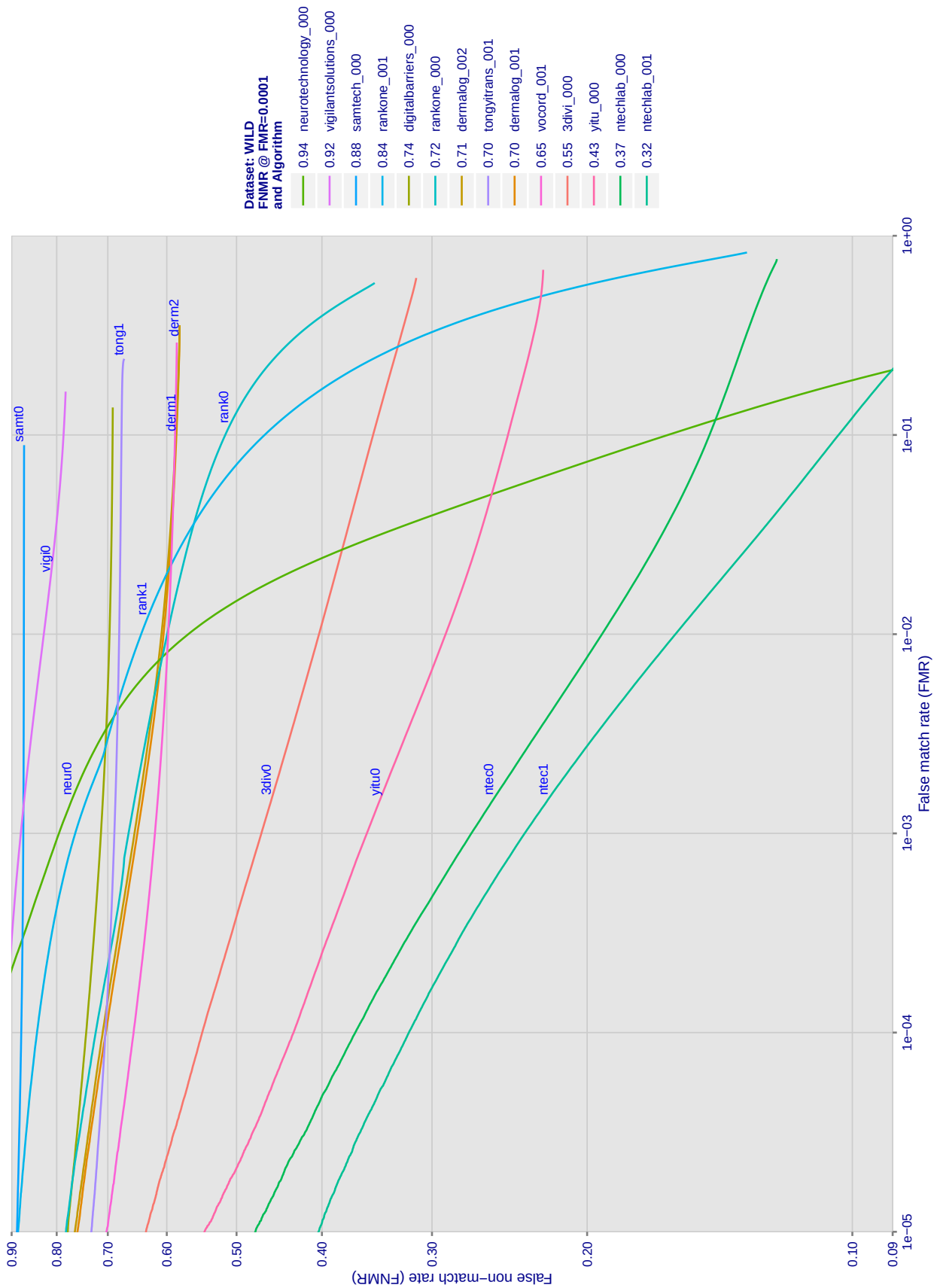


Figure 6: For the wild image comparisons, detection error tradeoff (DET) characteristics showing false non-match rate vs. false match rate plotted parametrically on threshold, T . The scales are logarithmic in order to show several decades of FMR.

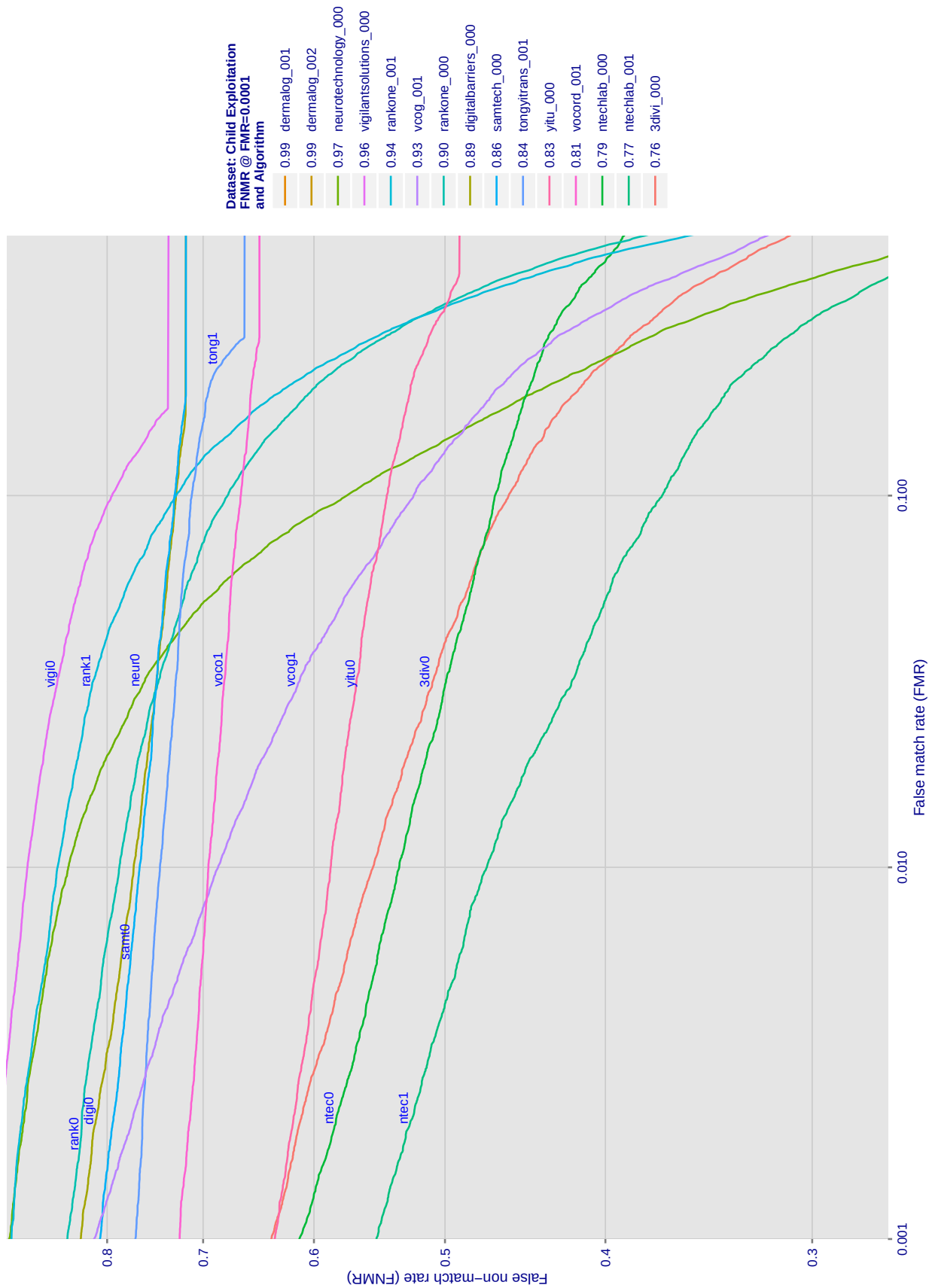


Figure 7: For child exploitation images, detection error tradeoff (DET) characteristics showing false non-match rate vs. false match rate plotted parametrically on threshold, T . The scales are logarithmic in order to show many decades of FMR. Accuracy is poor because many images have adverse quality characteristics, and because detection and enrollment fails.

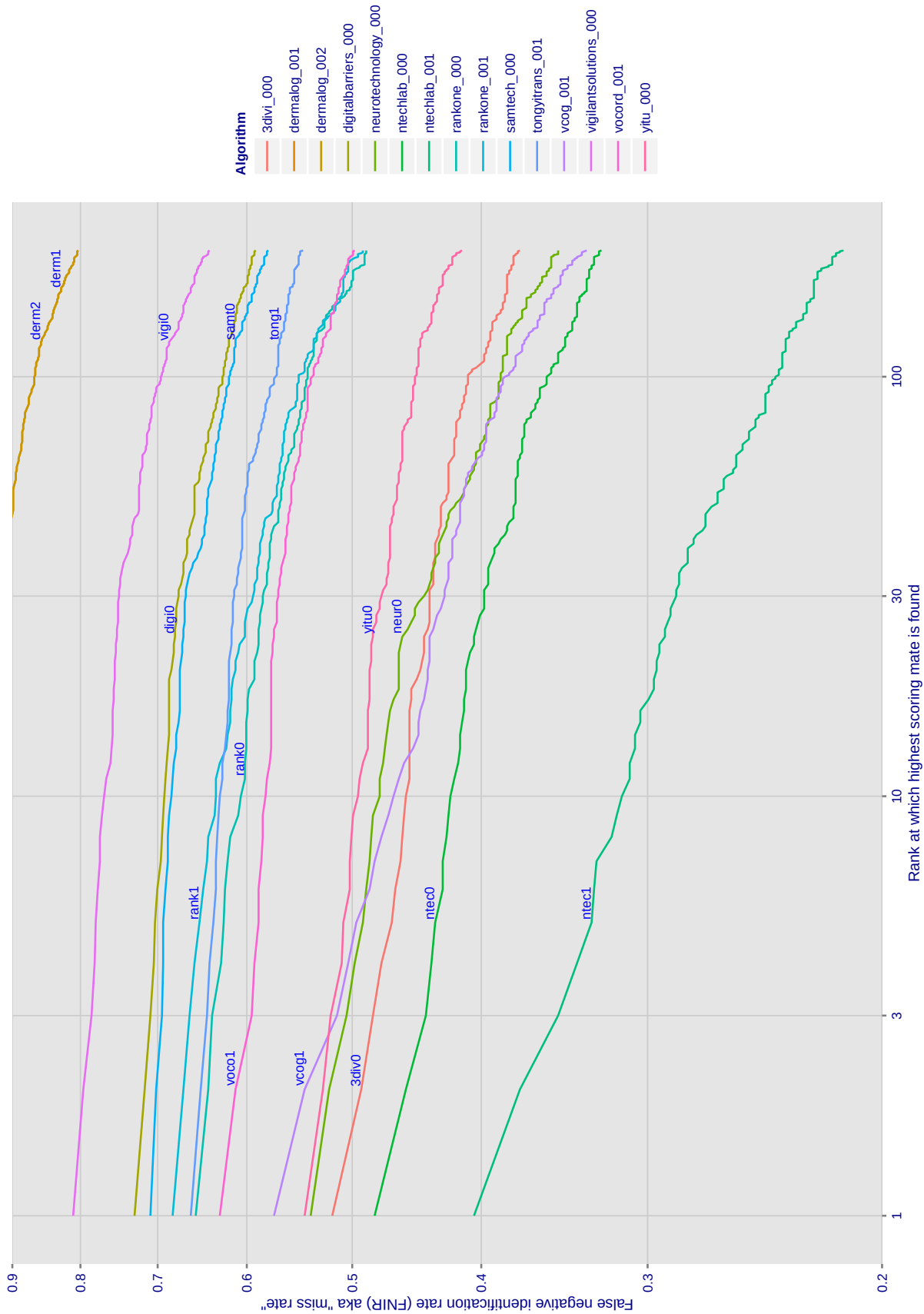


Figure 8: For child exploitation images, cumulative match characteristics (CMC) showing false negative identification rate vs. rank. This is simulation of a one-to-many search experiment - see discussion in section 4.2. The scales are logarithmic in order to show the effect of long candidate lists. Accuracy is poor but much improved relative to the 1:1 DETs of Fig. 7 because a search can succeed if any of a subject's several enrolled images matches the search image with a high score.

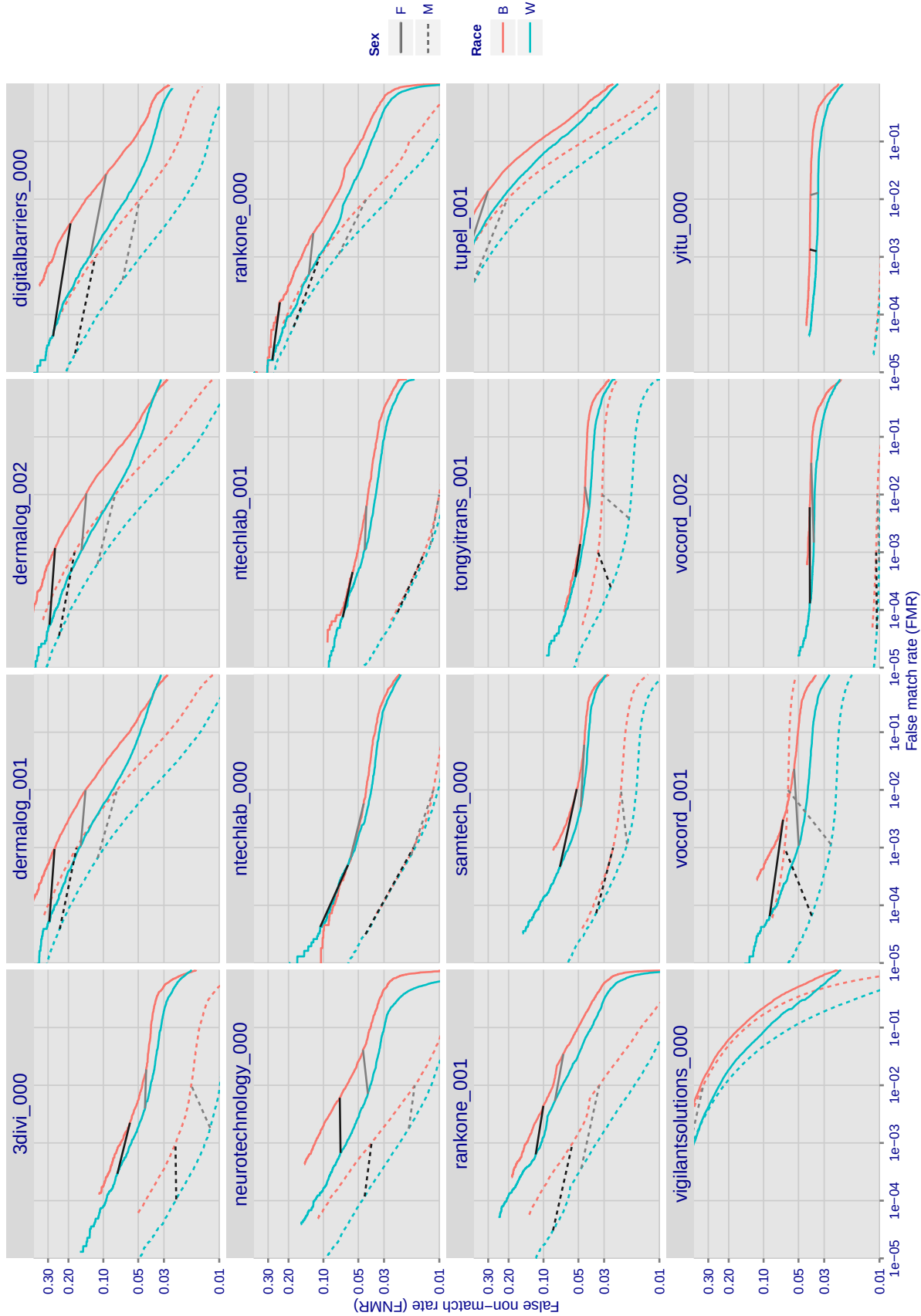


Figure 9: For the mugshot images, error tradeoff characteristics for white females, black females, black males and white males. The grey lines correspond to fixed thresholds, showing how both FNMR and FMR vary at one operating threshold. Important: Many of the plots will naively be read as saying whites gives lower error rates than blacks because the blue traces lie beneath the red ones. However, this is misleading and incomplete: The grey lines show the traces are generally shifted horizontally. Thus for the dermalog-001 algorithm FNMR for whites is higher than for blacks at a fixed threshold but, at the same time, FMR is higher for blacks - see Figure 14. As access control systems almost always operate at a fixed threshold, the naive interpretation is incorrect.

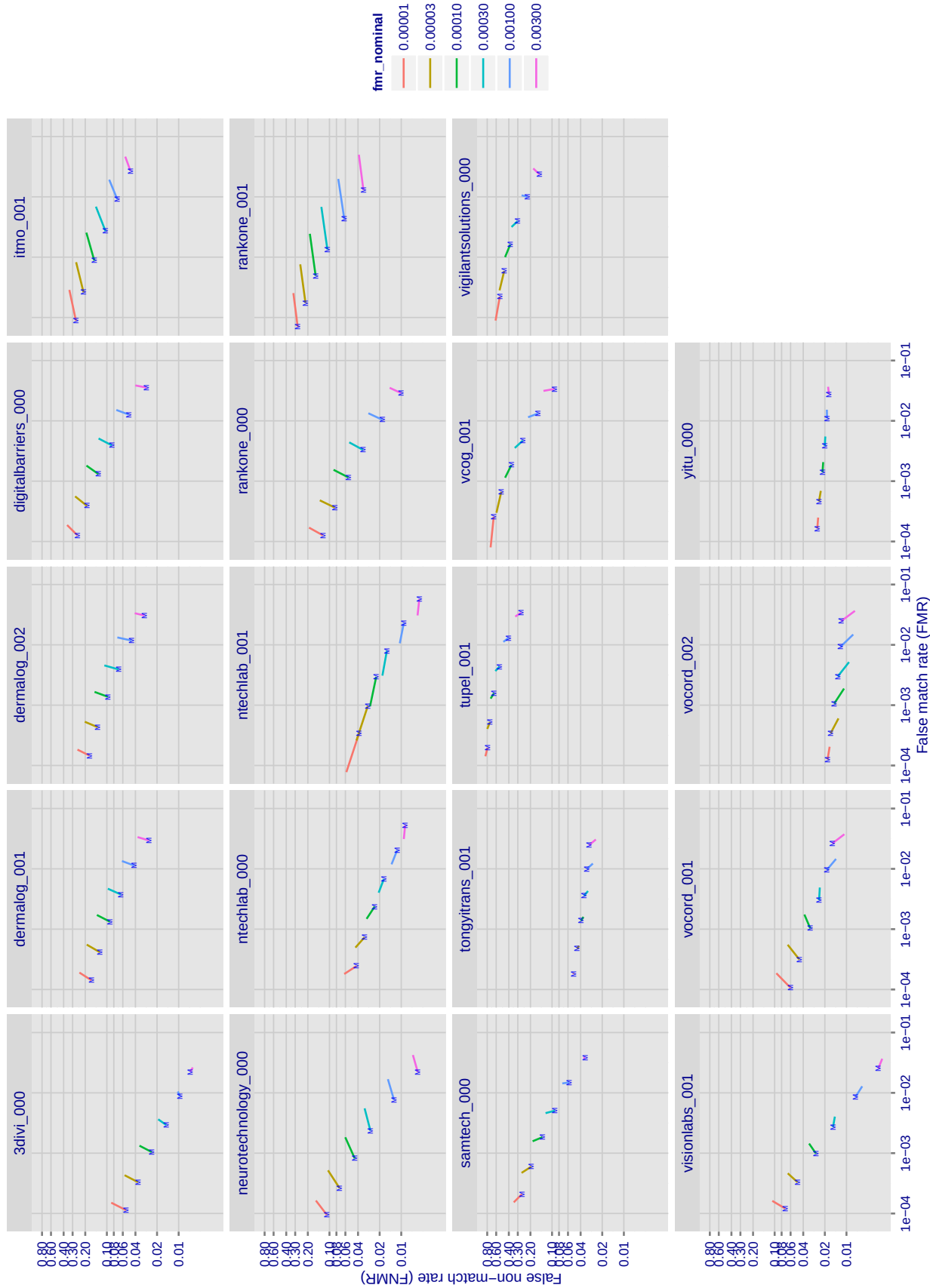


Figure 10: For the visa images, FNMR and FMR at six operating points along the DET characteristic. At each point a line is drawn between $(FMR, FNMR)_{\text{MALE}}$ and $(FMR, FNMR)_{\text{FEMALE}}$ showing how which sex has lower FMR and/or FNMR. The "M" label denotes male, the other end of the line corresponds to female. The six operating thresholds are selected to give the nominal false match rates given in the legend, and are computed over all impostor pairs regardless of age, sex, and place of birth. The plotted FMR values are broadly an order of magnitude larger than the nominal rates because FMR is computed over demographically-matched impostor pairs i.e individuals of the same sex, from the same geographic region (see section 4.6.1), and the same age group (see section 4.6.2).

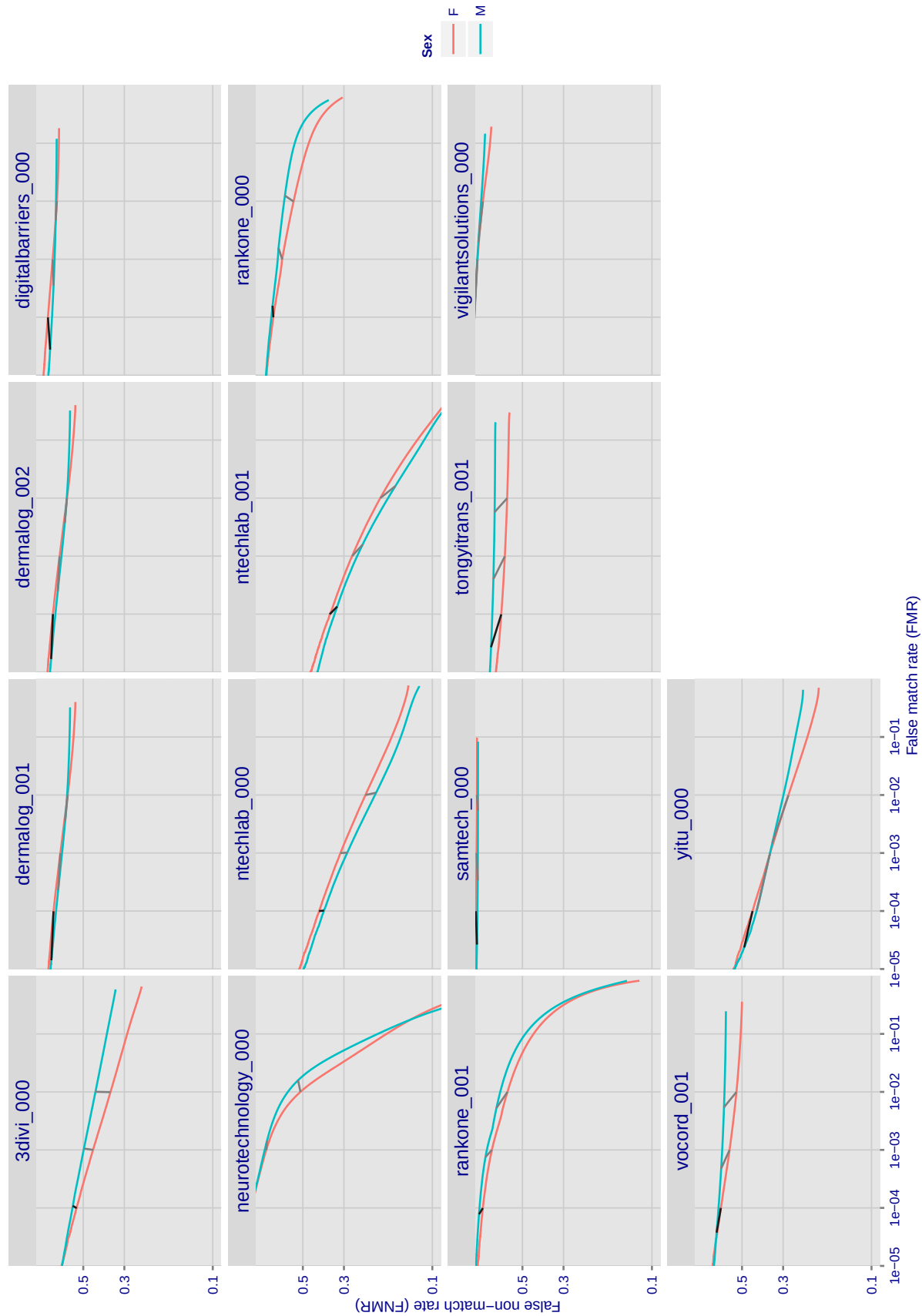


Figure 11: For the wild image comparisons, detection error tradeoff (DET) characteristics showing false non-match rate vs. false match rate plotted parametrically on threshold, T. Error rates are higher here than in the generic wild DET (Fig 6) because the impostor pairs here are same-sex only. The scales are logarithmic in order to show several decades of FMR.

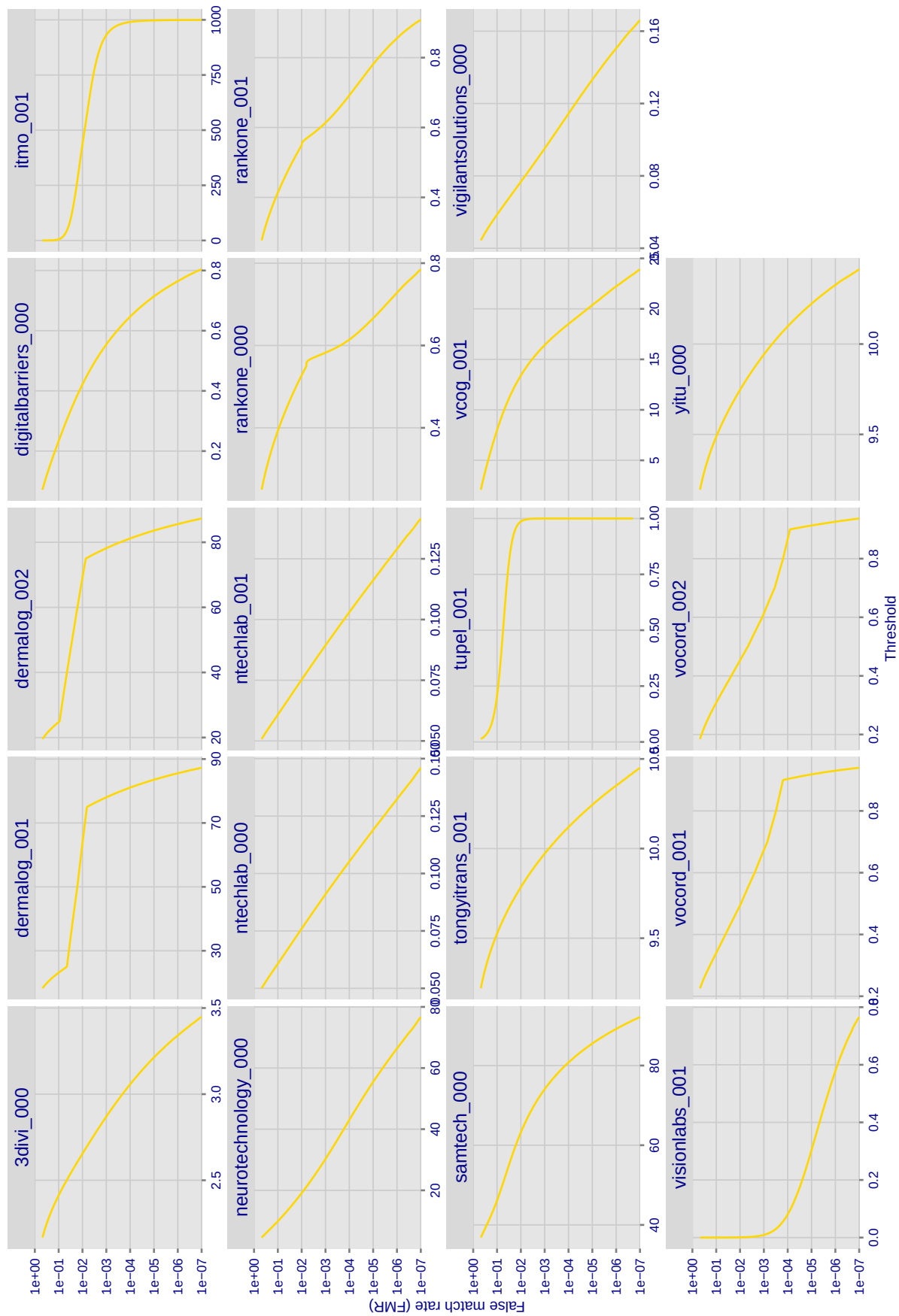


Figure 12: For the visa images, the false match calibration curves show false match rate vs. threshold. These curves apply to zero-effort impostors. As shown later (sec. 4.6), FMR is higher for demographic-matched impostors.

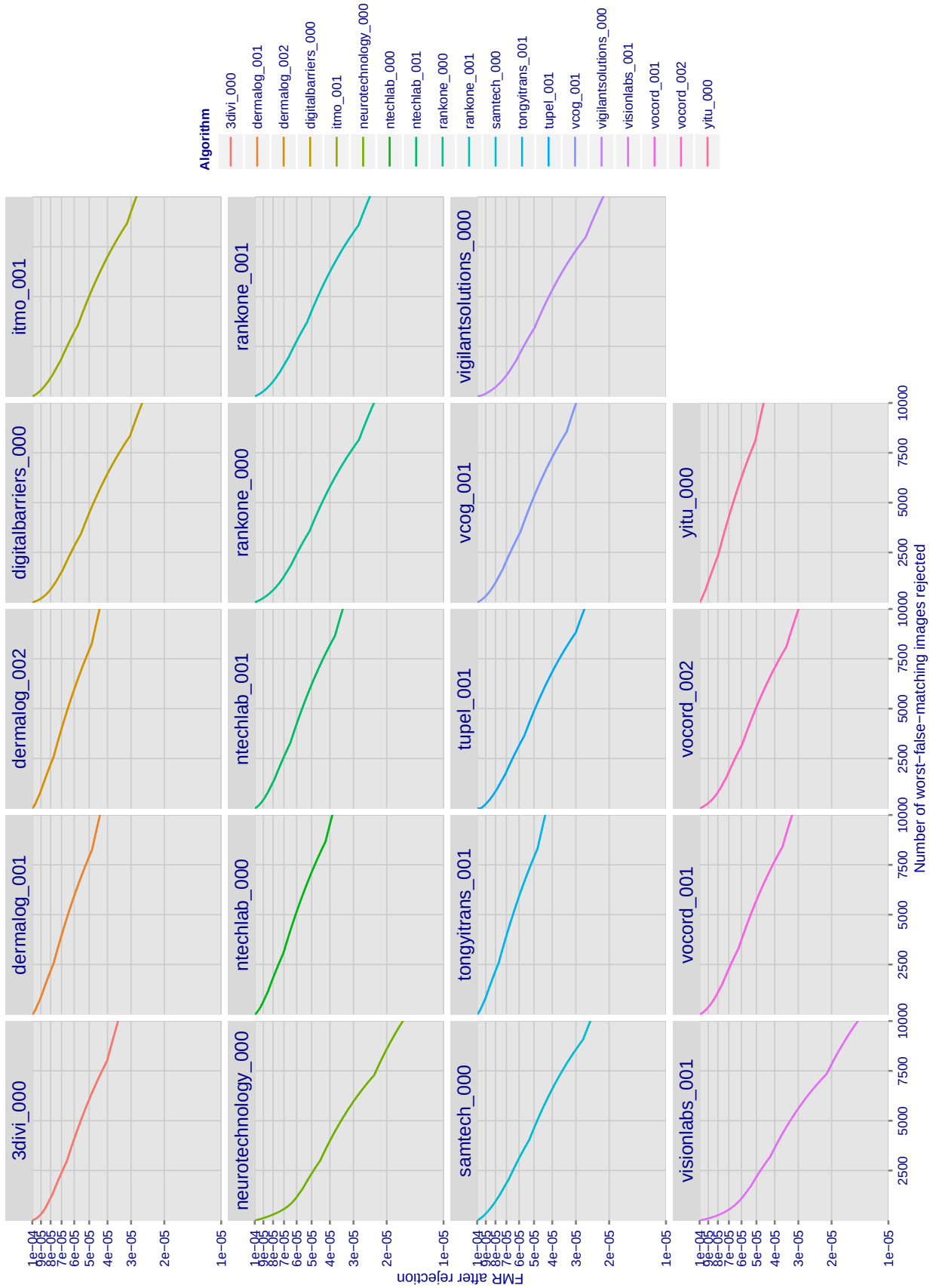


Figure 13: For the visa images, the curves show how false matches are concentrated in certain images. Specifically each line plots $FMR(k)$ with k the number of images rejected in decreasing order of how many false matches that image was involved in. $FMR(0) = 10^{-4}$. In terms of the biometric zoo, the most “wolf-ish” images are rejected first i.e. those enrollment or verification images involved in false matches. A flat response is considered superior. A steeply descending response indicates that certain kinds of images false match against others. A hypothetical example would be if images of men with particular mustaches falsely match others.

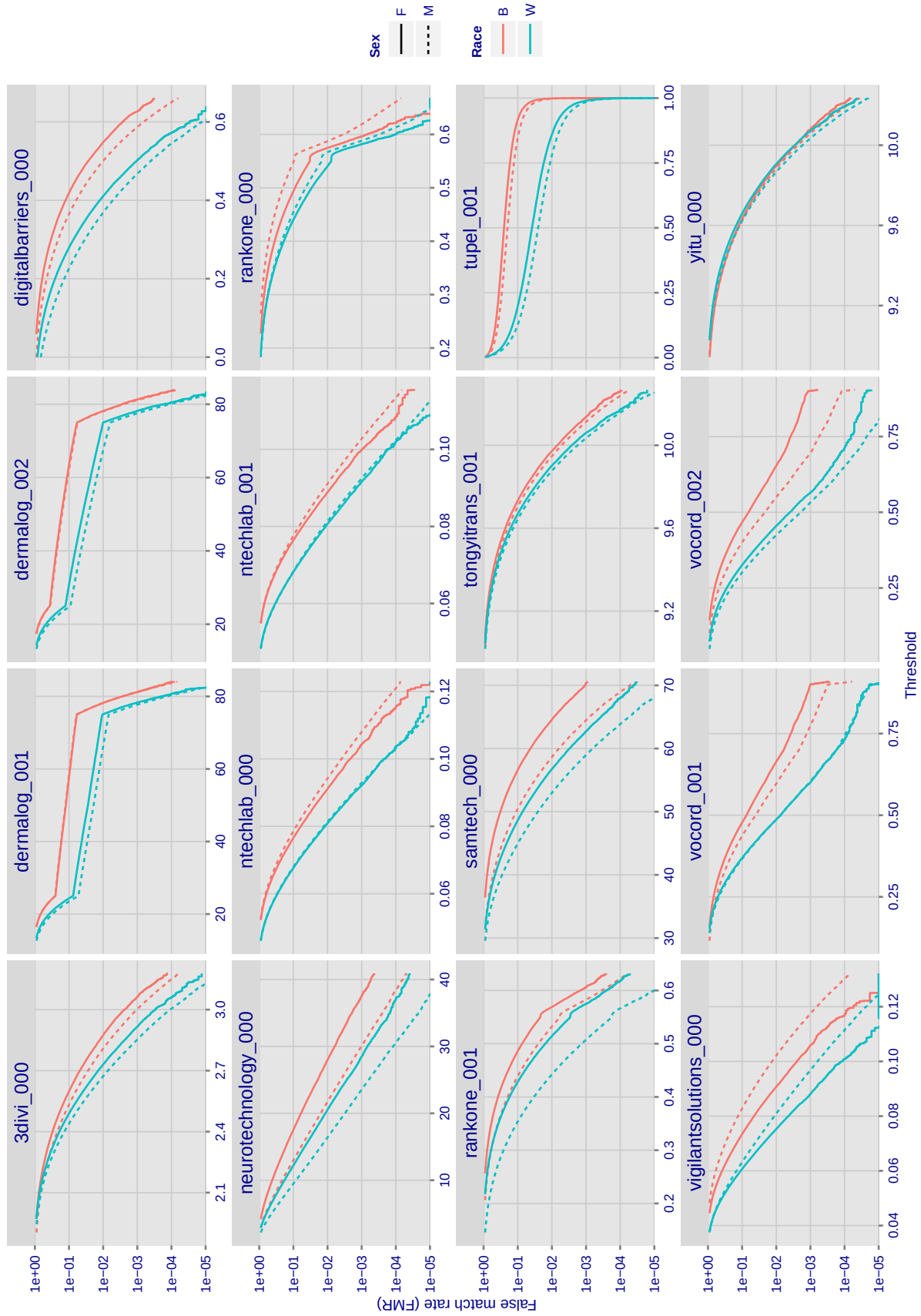


Figure 14: For the mugshot images, the false match calibration curves show false match rate vs. threshold. Separate curves appear for white females, black females, black males and white males.

4.5 Genuine distribution stability

4.5.1 Effect of birth place on the genuine distribution

Background: Both skin tone and bone structure vary geographically. Prior studies have reported variations in FNMR and FMR.

Goal: To measure false non-match rate (FNMR) variation with country of birth.

Methods: Thresholds are determined that give $FMR = \{0.001, 0.0001\}$ 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 15 shows FNMR by country of birth for the two thresholds.

Caveats: The results may not relate to subject-specific properties. Instead they could reflect image-specific quality differences, which could occur due to collection protocol or software processing variations.

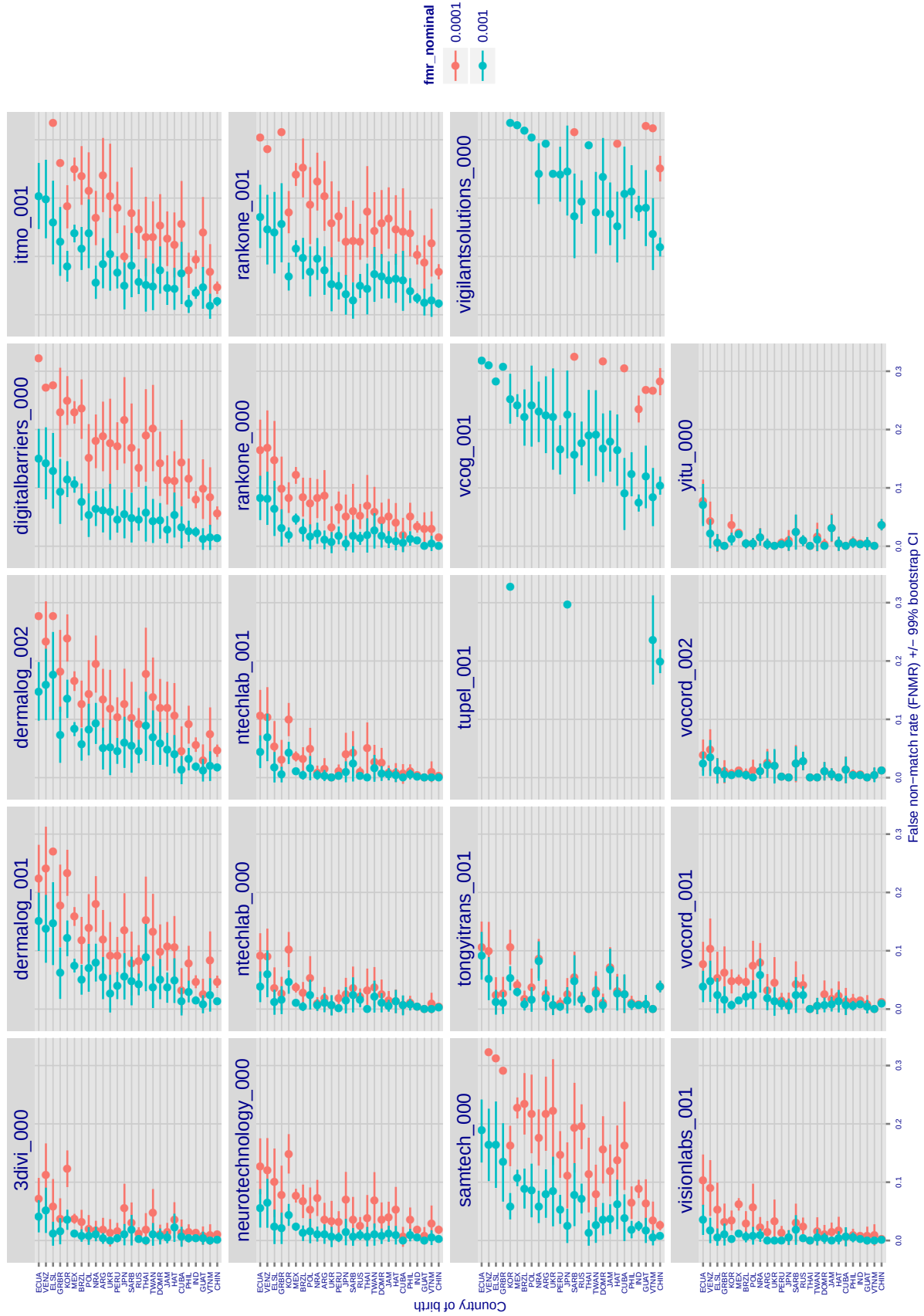


Figure 15: For the visa images, the dots show FNMR by country of birth for two operating thresholds corresponding to $FMR = \{0.001, 0.0001\}$ computed over all $O(10^{10})$ impostor scores. The figures shows an order of magnitude variation in FNMR across country of birth; these effects are due to quality variations. The least accurate countries vary by algorithm.

4.5.2 Effect of age on genuine subjects

Background: Faces change appearance throughout life. Face recognition algorithms have previously been 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: Using the visa images, 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 30 individuals are included in the analysis.

Results: For the visa images, Figure 16 shows how false non-match rates for genuine users, as a function of age group.

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, $AGE > 72$, contains fewer than 140 mated pairs, and may be affected by small sample size.

Caveats: None.

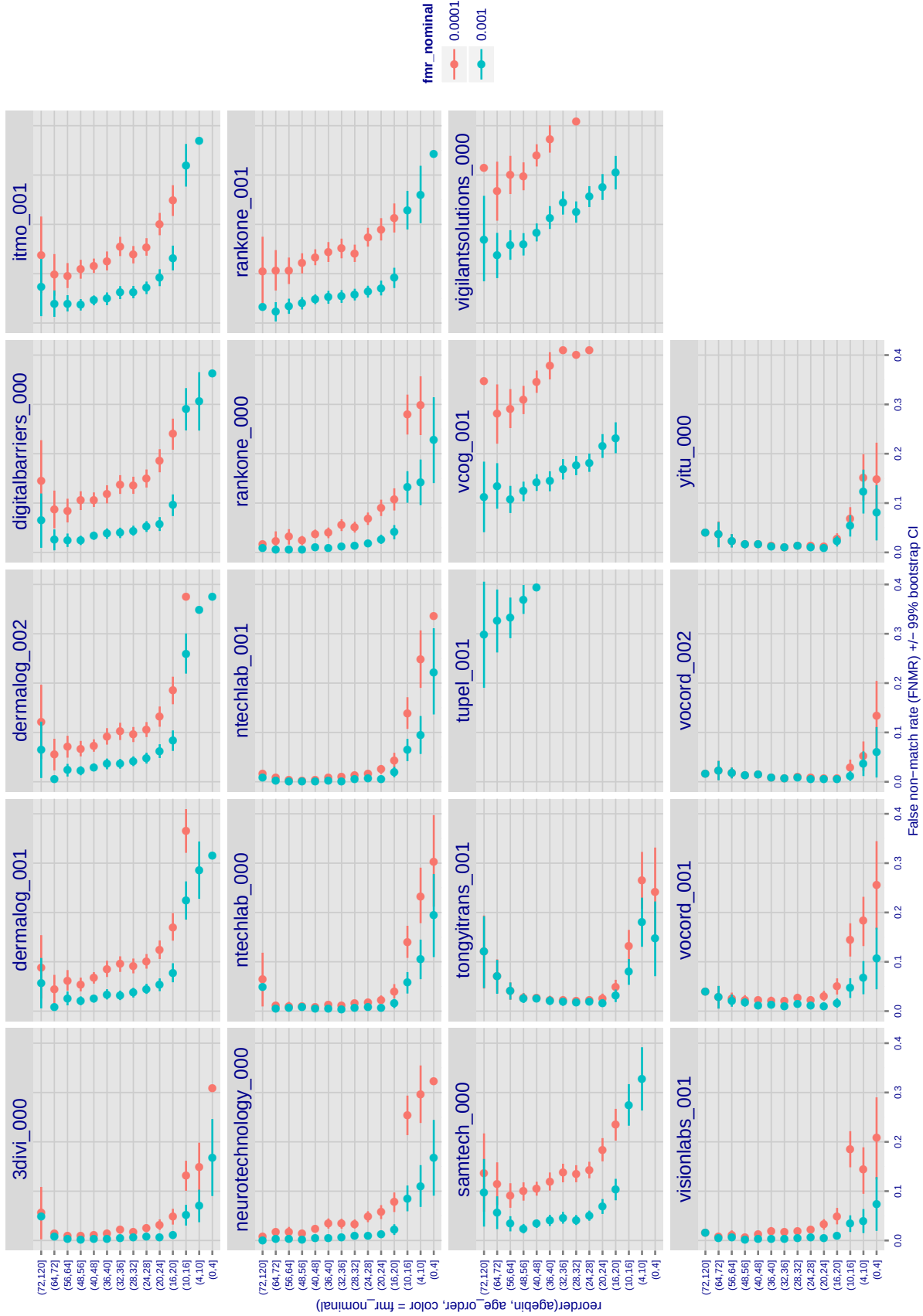


Figure 16: For the visa images, the dots show FNMR by age group for two operating thresholds corresponding to $FMR = \{0.001, 0.0001\}$ computed over all $O(10^{10})$ impostor scores. Given a pair of face images taken at different times, we assign a false non-match to the bin that is the arithmetic average 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.

4.6 Impostor distribution stability

4.6.1 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 birth region of the impostor and enrollee on false match rates.
- ▷ To determine whether some algorithms give better impostor distribution stability.

Methods:

- ▷ For the visa images, 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.
- ▷ For the visa images, 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 .

$$\text{FMR}(r_1, r_2, T) = \frac{\sum_{i=1}^{N_{r_1, r_2}} H(s_i - T)}{N_{r_1, r_2}} \quad (5)$$

where the same threshold, T , is used in all cells, and H is the unit step function. The threshold is set to give $\text{FMR}(T) = 0.001$ 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 countries¹ 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 51.

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:

- ▷ The on-diagonal elements correspond to within-region impostors. FMR is generally above the nominal value of $\text{FMR} = 0.001$. Particularly there is usually higher FMR in, Sub-Saharan Africa, South Asia, and the Caribbean. Europe and Central Asia, on the other hand, usually give FMR closer to the nominal value.
- ▷ The off-diagonal elements correspond to across-region impostors. The highest FMR is produced between the Caribbean and Sub-Saharan Africa.
- ▷ Algorithms vary.

¹These are Argentina, Australia, Brazil, Chile, China, Costa Rica, Cuba, Czech Republic, Dominican Republic, Ecuador, Egypt, El Salvador, Germany, Ghana, Great Britain, Greece, Guatemala, 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.

- ▷ We computed the same quantities for a global FMR = 0.0001. The effects are similar.

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.

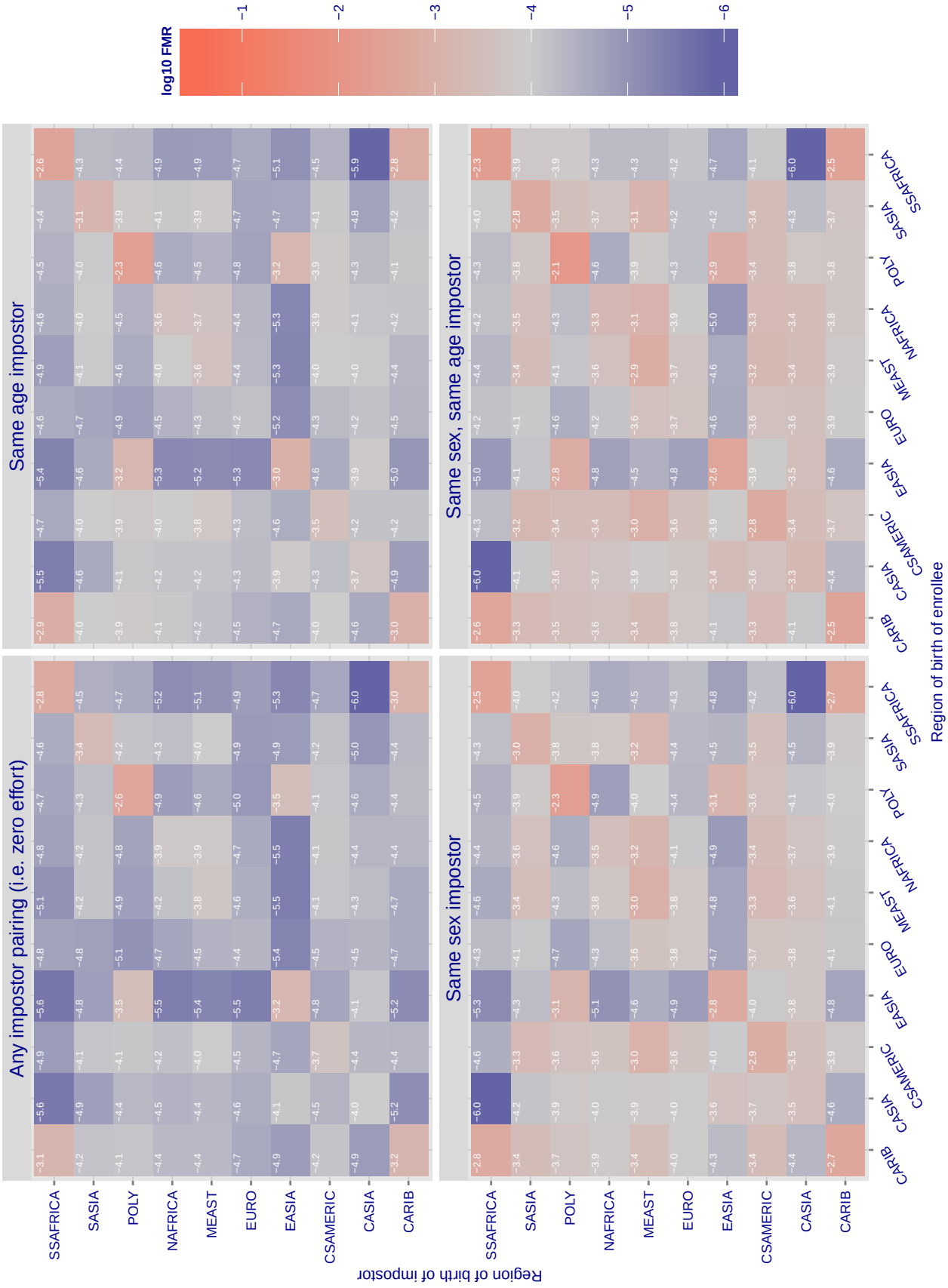
Cross region FMR at threshold $T = 3.057$ for algorithm 3divi_000, giving $FMR(T) = 0.0001$ globally.

Figure 17: For algorithm 3divi-000 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

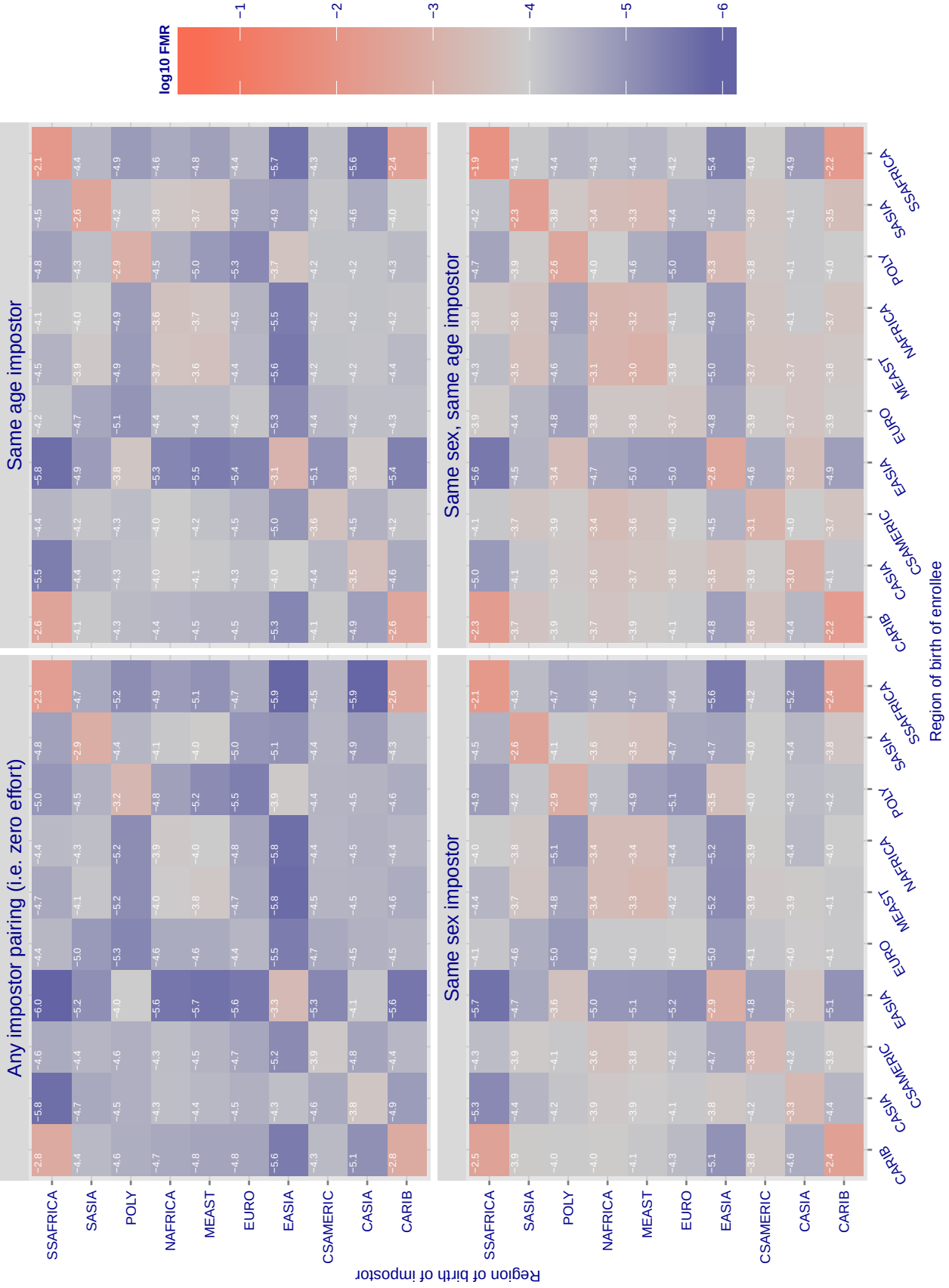
Cross region FMR at threshold $T = 81.064$ for algorithm dermalog_001, giving $FMR(T) = 0.0001$ globally.

Figure 18: For algorithm dermalog-001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

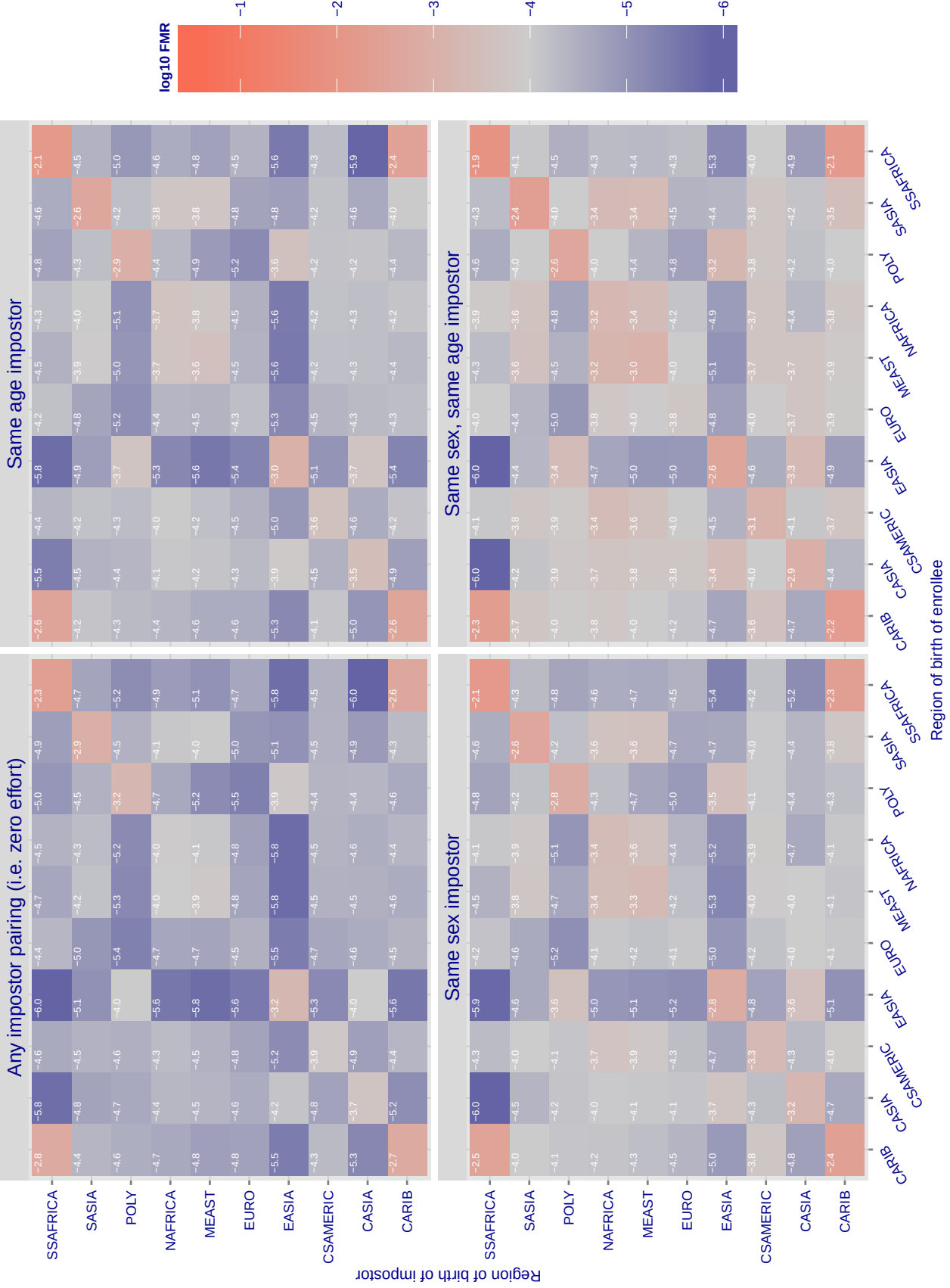
Cross region FMR at threshold $T = 81.164$ for algorithm dermalog_002, giving $FMR(T) = 0.0001$ globally.

Figure 19: For algorithm dermalog-002 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

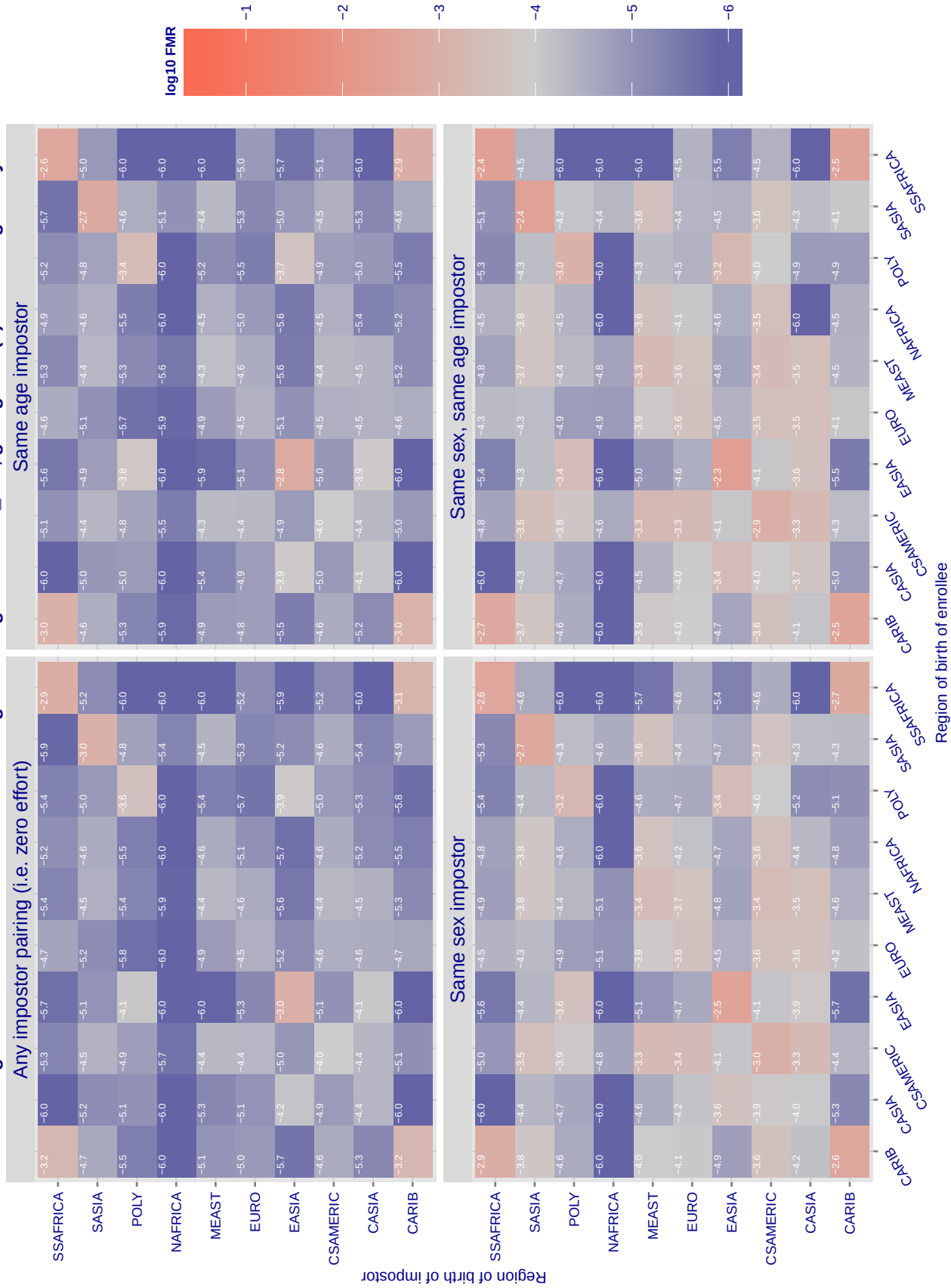
Cross region FMR at threshold $T = 0.646$ for algorithm digitalbarriers_000, giving $FMR(T) = 0.0001$ globally.

Figure 20: For algorithm digitalbarriers-000 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

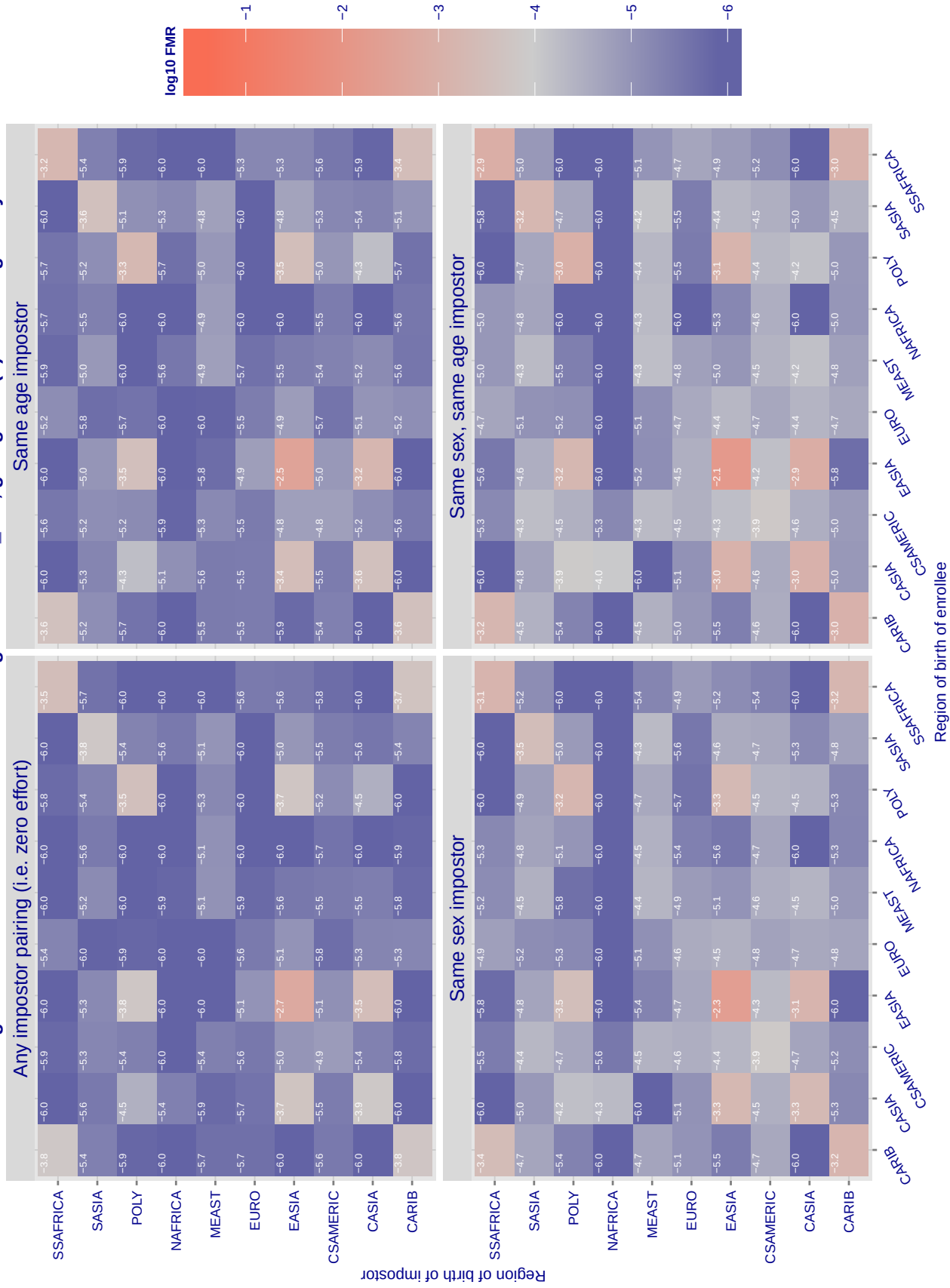
Cross region FMR at threshold $T = 990.194$ for algorithm itmo_001, giving $FMR(T) = 0.0001$ globally.

Figure 21: For algorithm itmo-001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

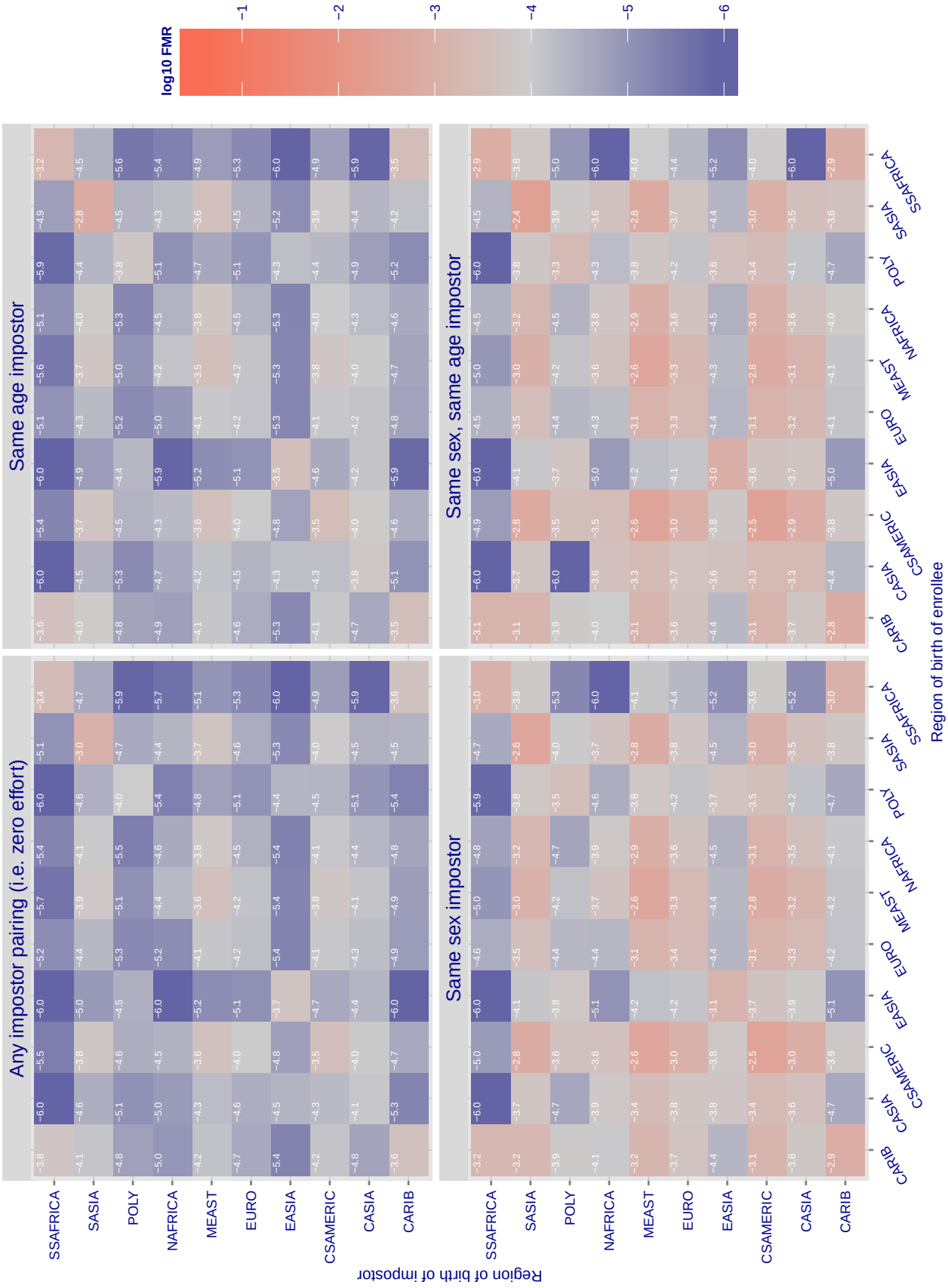
Cross region FMR at threshold $T = 43.010$ for algorithm neurotechnology_000, giving $FMR(T) = 0.0001$ globally.

Figure 22: For algorithm neurotechnology-000 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

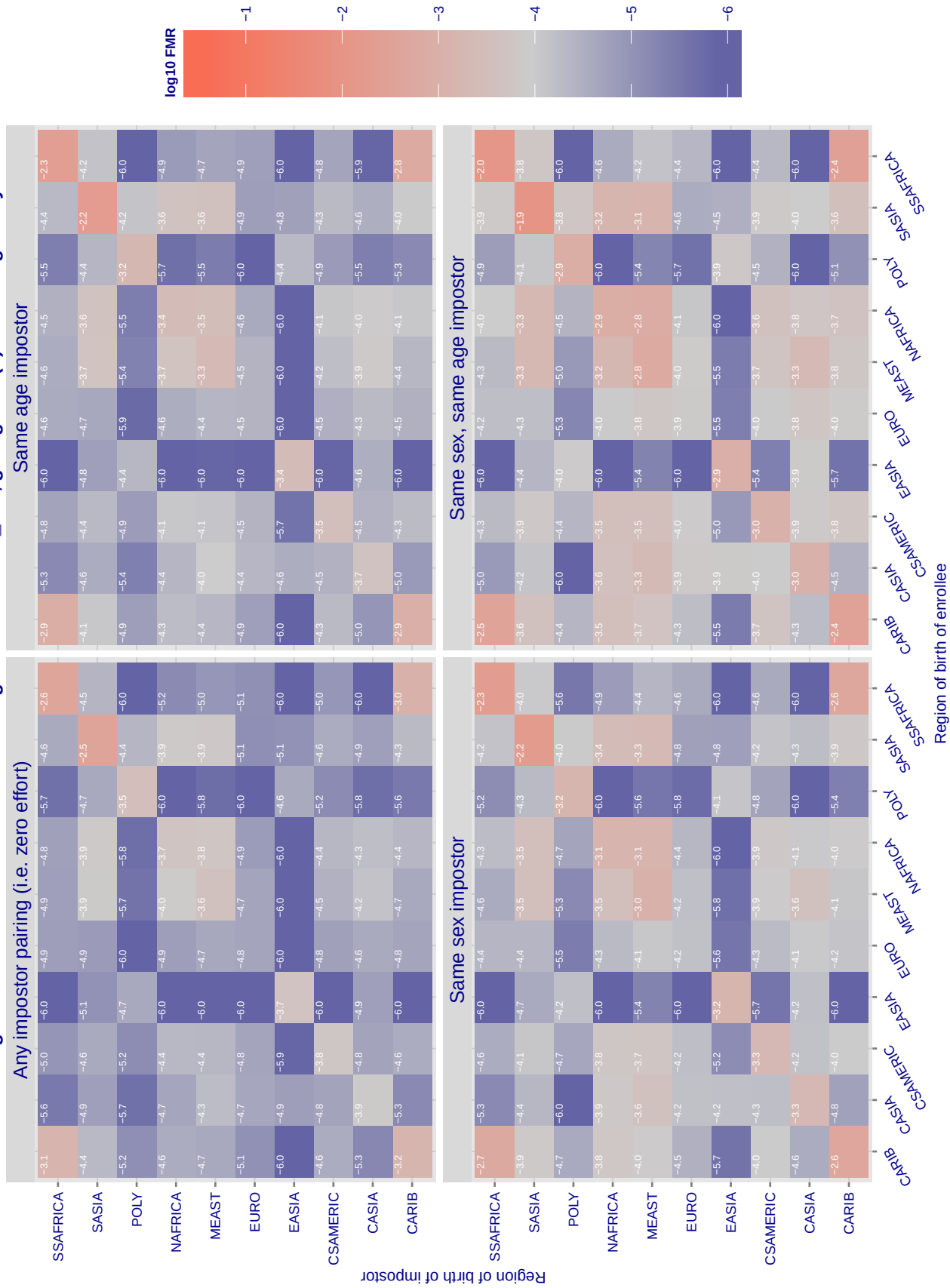
Cross region FMR at threshold $T = 0.105$ for algorithm ntechlab_000, giving $FMR(T) = 0.0001$ globally.

Figure 23: For algorithm ntechlab-000 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

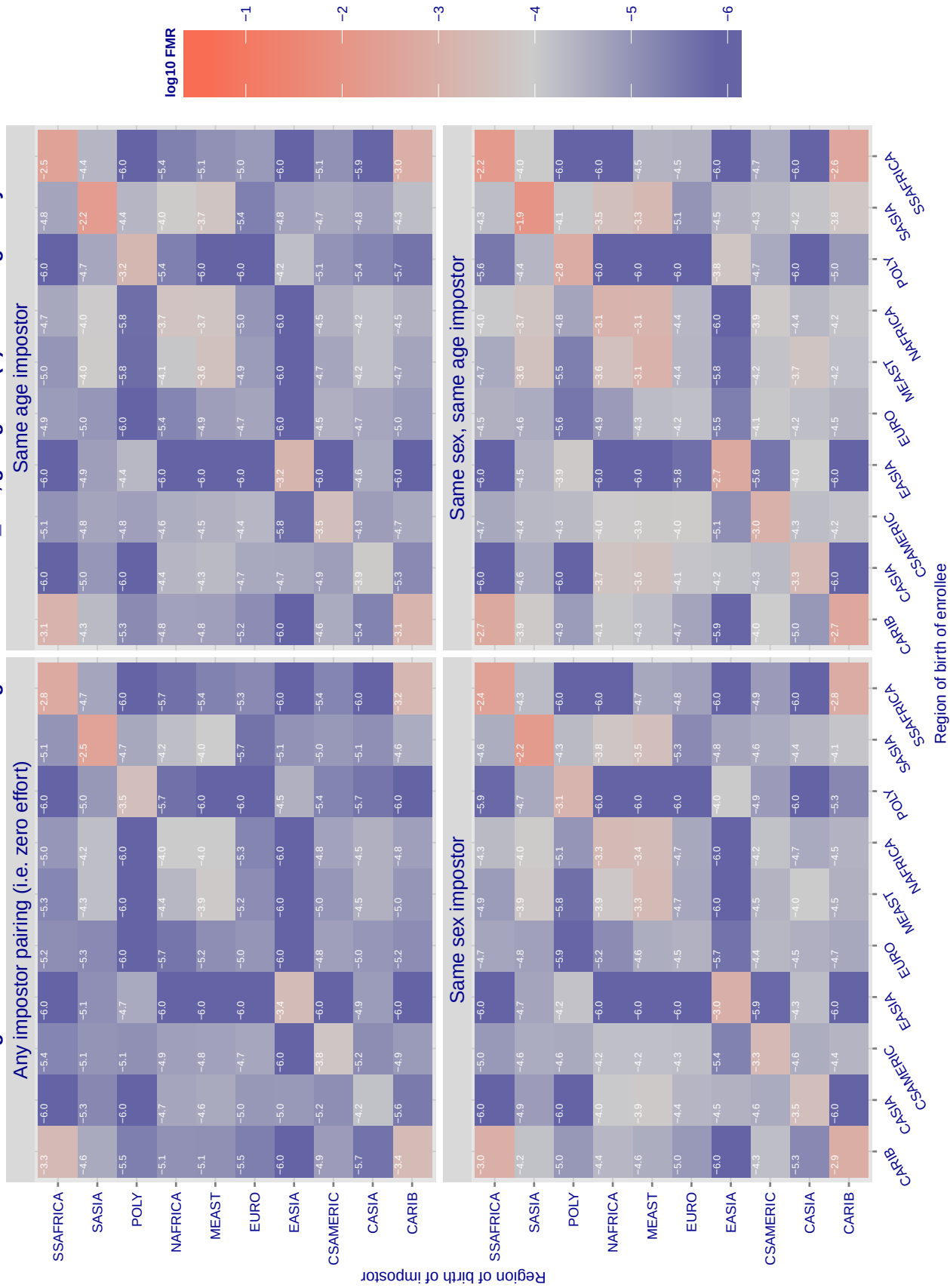
Cross region FMR at threshold $T = 0.103$ for algorithm ntechlab_001, giving $FMR(T) = 0.0001$ globally.

Figure 24: For algorithm ntechlab-001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

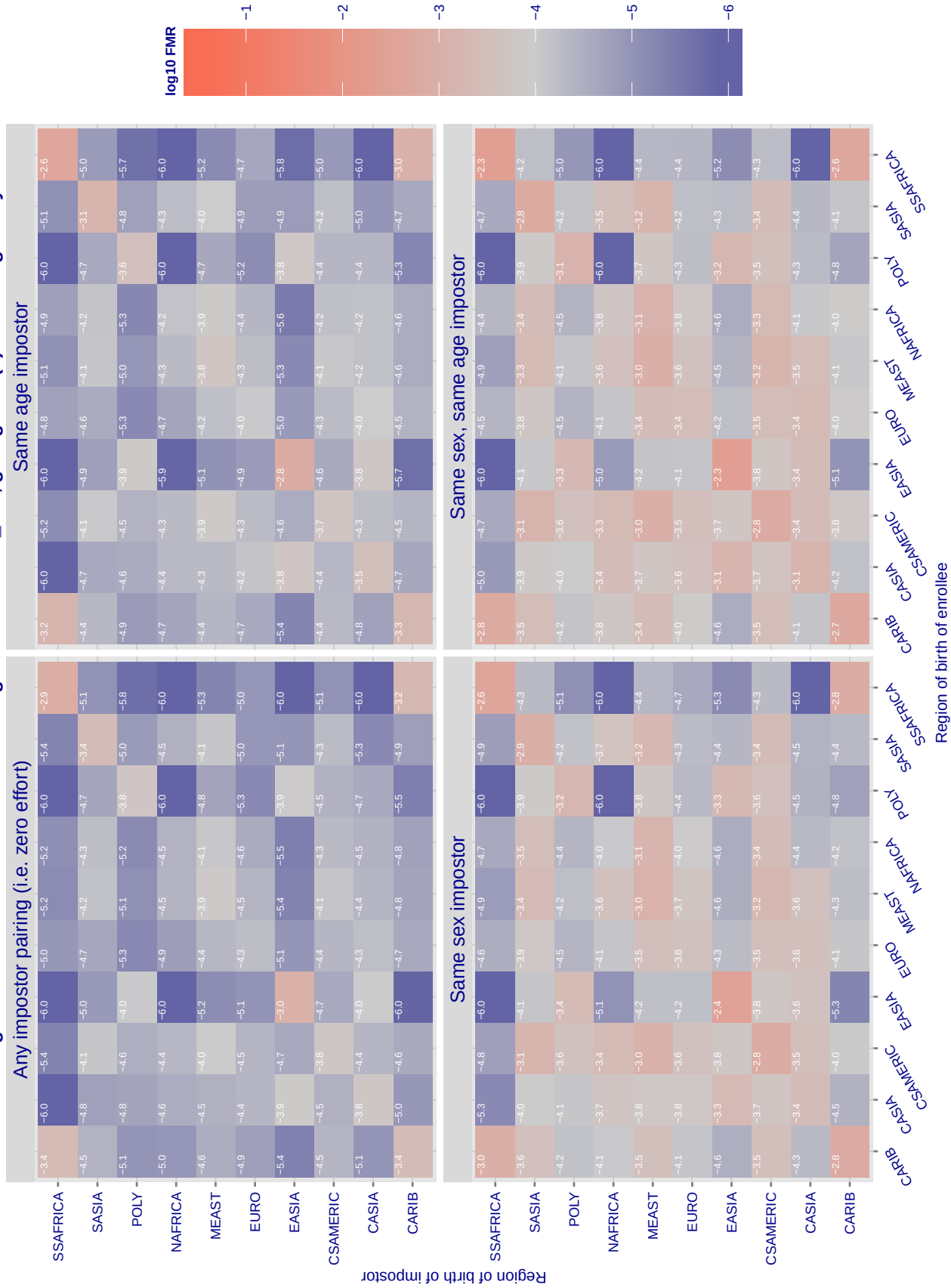
Cross region FMR at threshold $T = 0.614$ for algorithm rankone_000, giving $FMR(T) = 0.0001$ globally.

Figure 25: For algorithm rankone-000 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

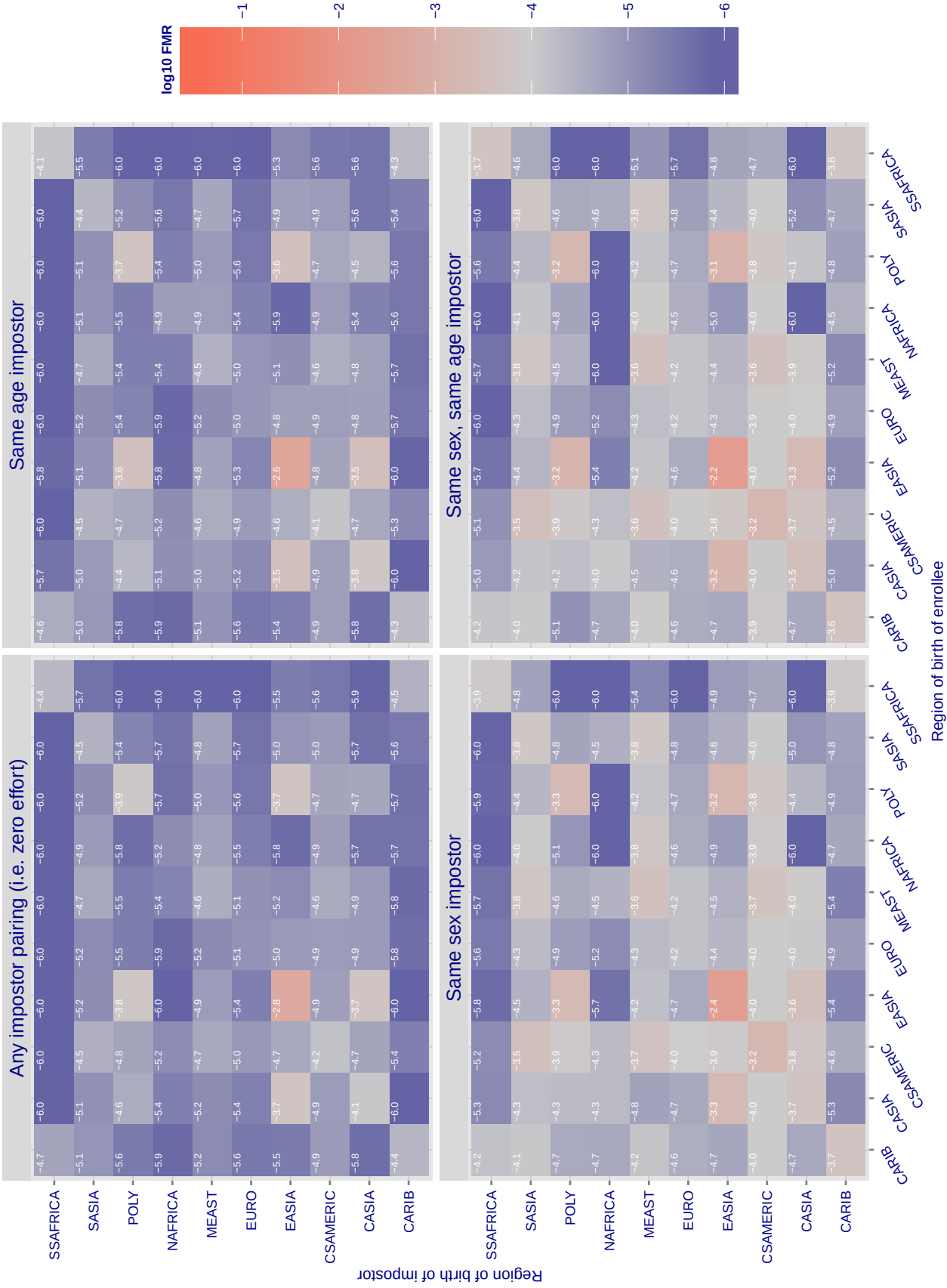
Cross region FMR at threshold $T = 0.692$ for algorithm rankone_001, giving $FMR(T) = 0.0001$ globally.

Figure 26: For algorithm rankone-001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

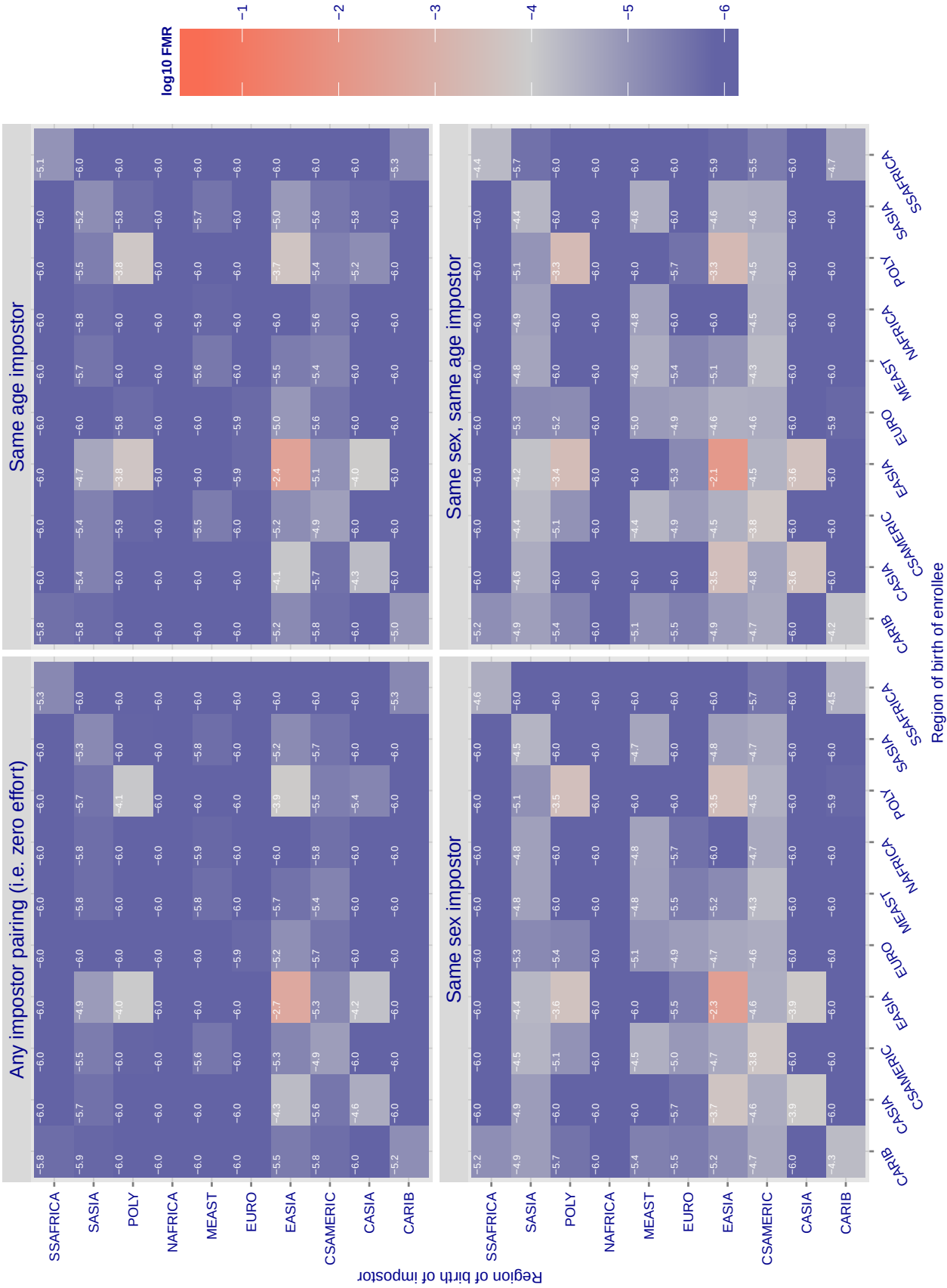
Cross region FMR at threshold $T = 80.766$ for algorithm samtech_000, giving $FMR(T) = 0.0001$ globally.

Figure 27: For algorithm samtech-000 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

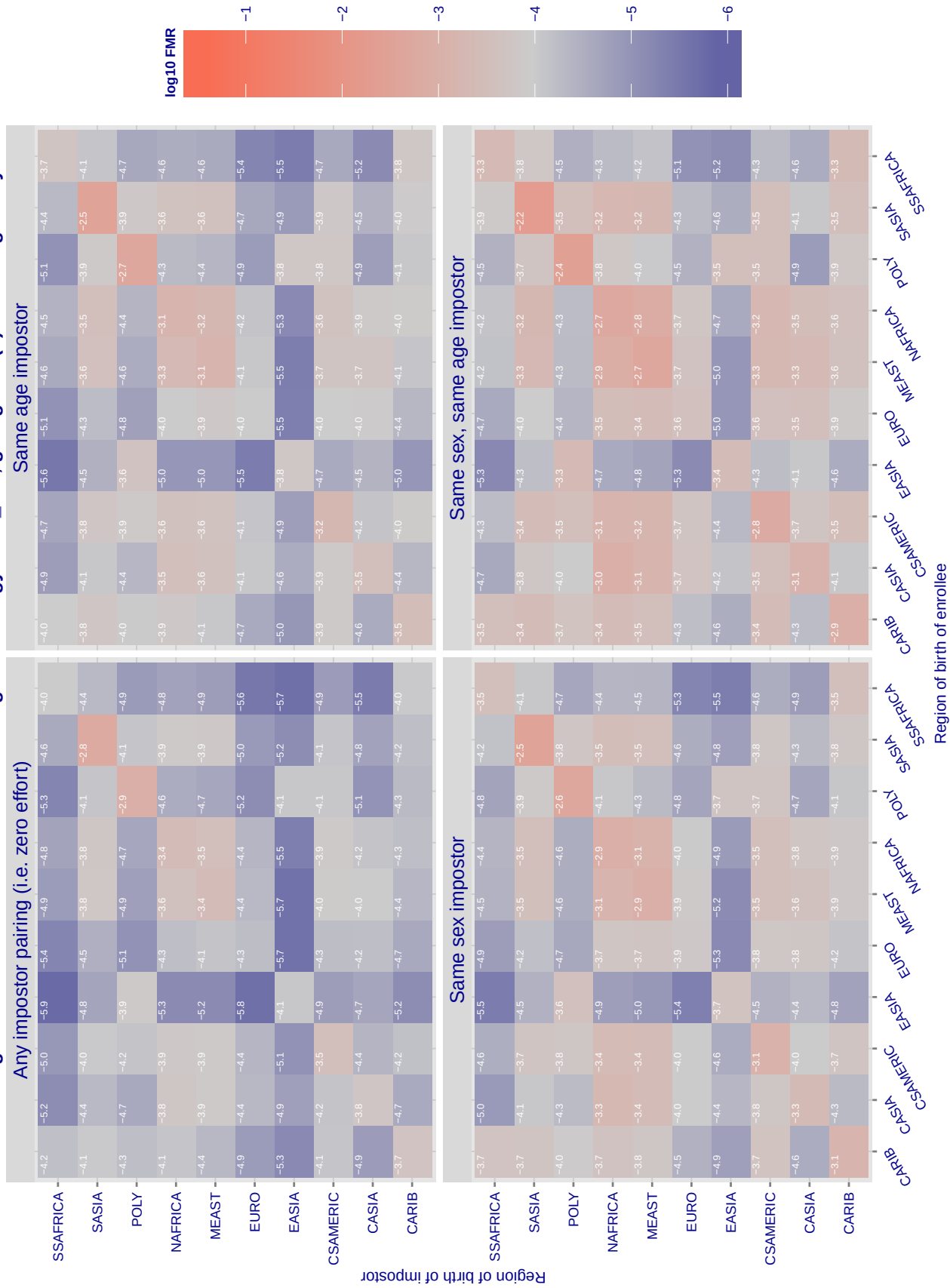
Cross region FMR at threshold $T = 10.120$ for algorithm tongyitrans_001, giving $FMR(T) = 0.0001$ globally.

Figure 28: For algorithm tongyitrans-001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

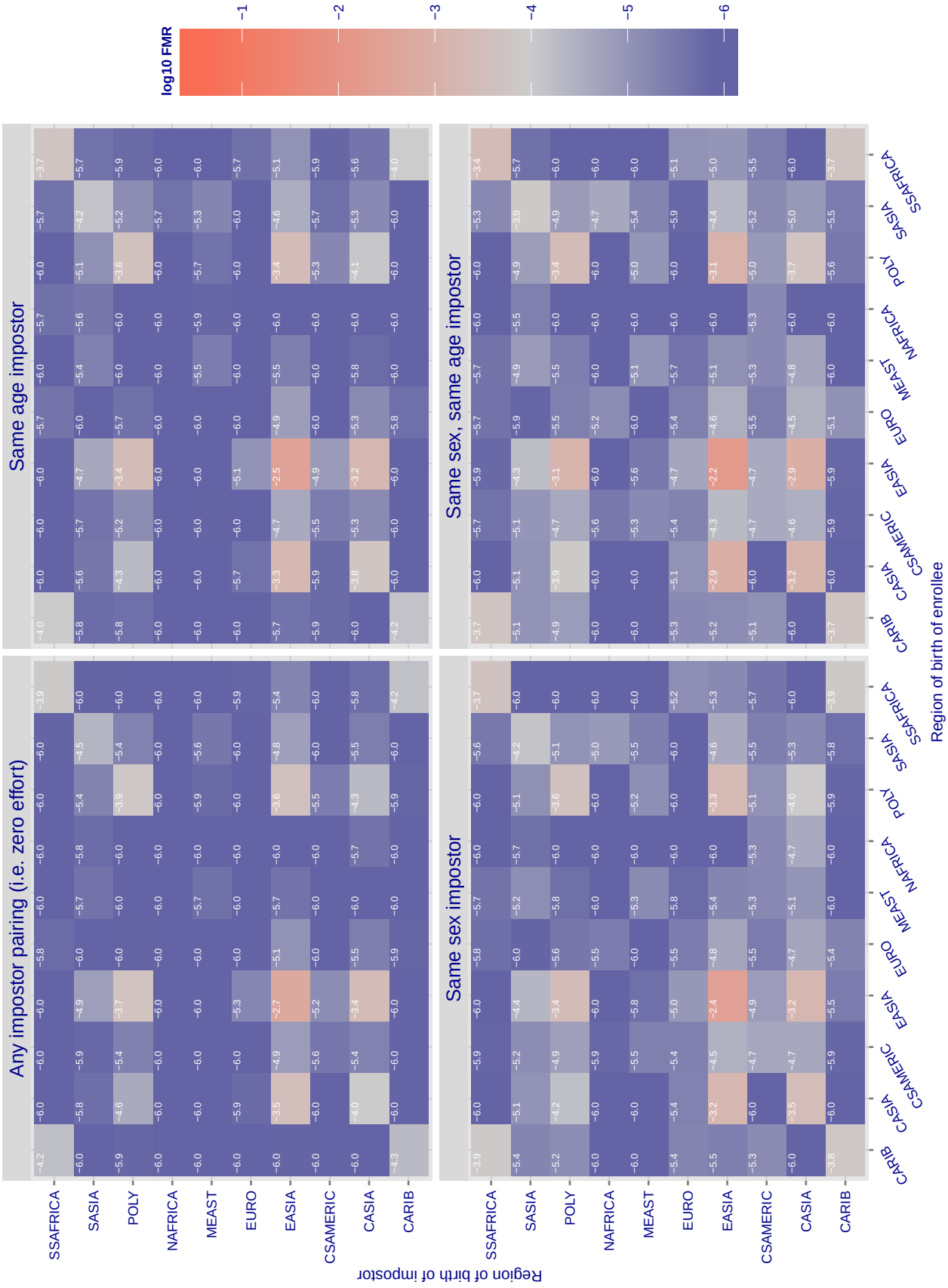
Cross region FMR at threshold $T = 1.000$ for algorithm tuple_001, giving $FMR(T) = 0.0001$ globally.

Figure 29: For algorithm tuple_001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

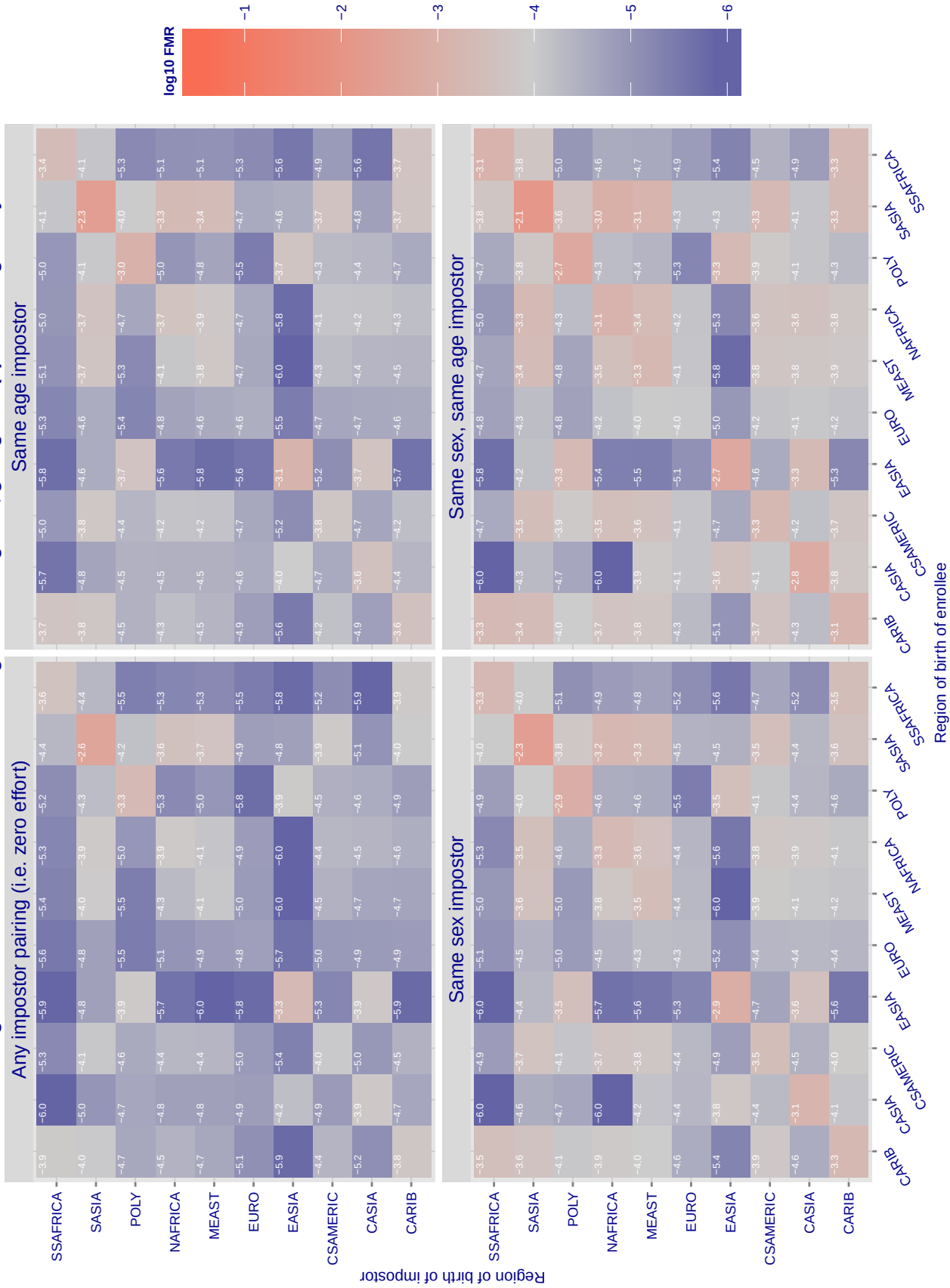
Cross region FMR at threshold $T = 18.505$ for algorithm vcog_001, giving $FMR(T) = 0.0001$ globally.

Figure 30: For algorithm vcog-001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

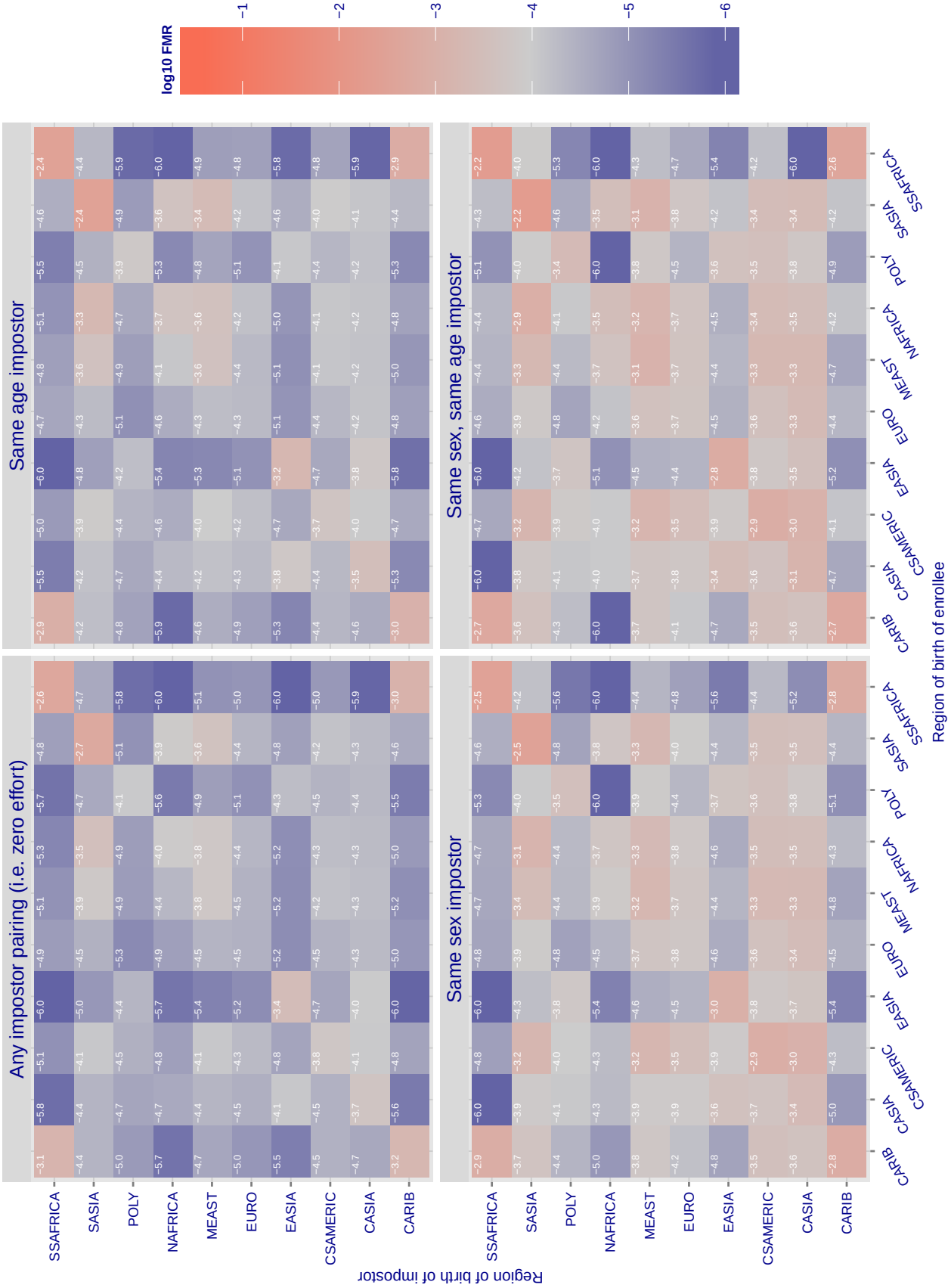
Cross region FMR at threshold $T = 0.114$ for algorithm vigilantsolutions_000, giving $FMR(T) = 0.0001$ globally.

Figure 31: For algorithm vigilantsolutions-000 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

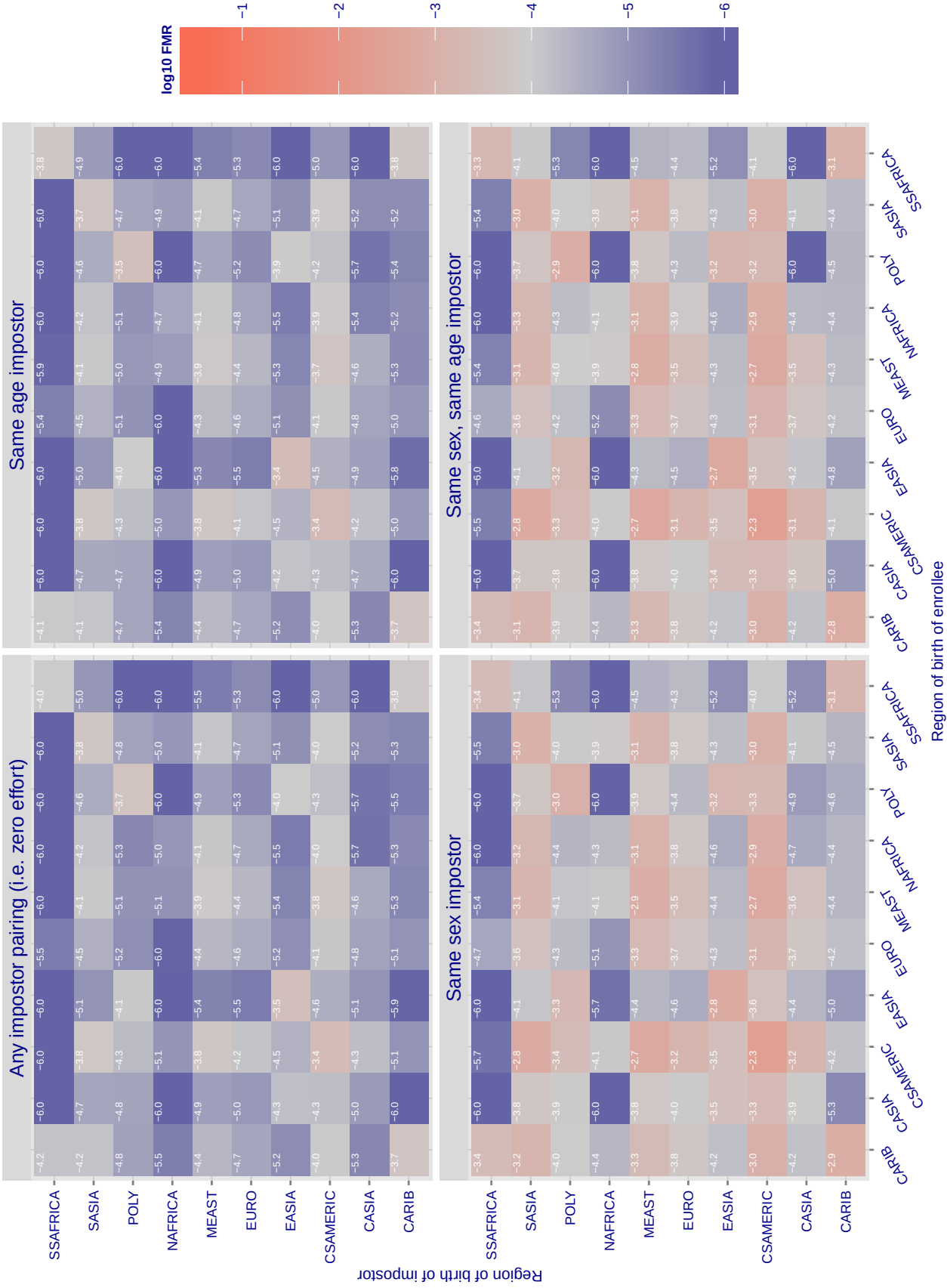
Cross region FMR at threshold $T = 0.080$ for algorithm visionlabs_001, giving $FMR(T) = 0.0001$ globally.

Figure 32: For algorithm visionlabs-001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

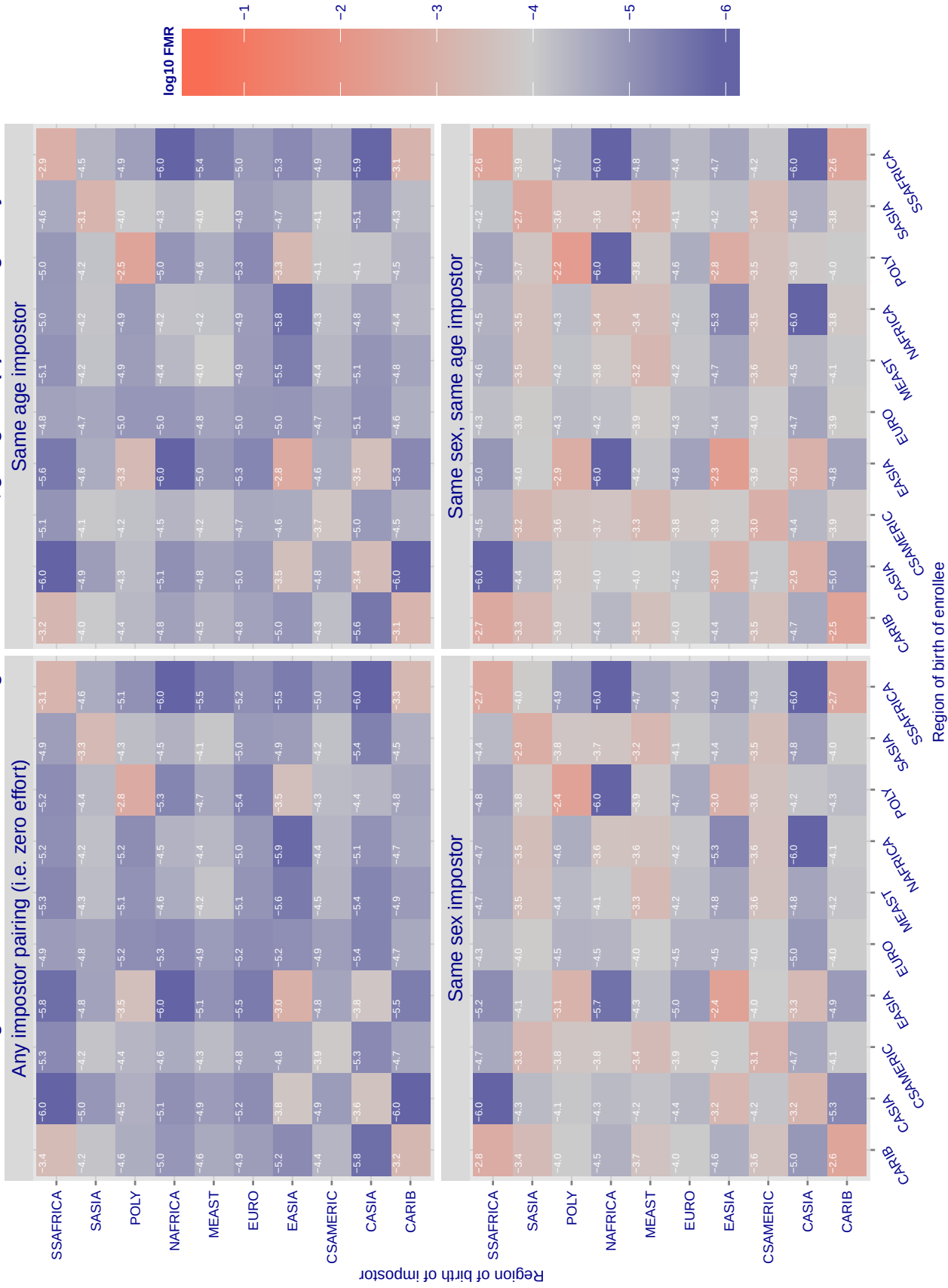
Cross region FMR at threshold $T = 0.903$ for algorithm vocord_001, giving $FMR(T) = 0.0001$ globally.

Figure 33: For algorithm vocord-001 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

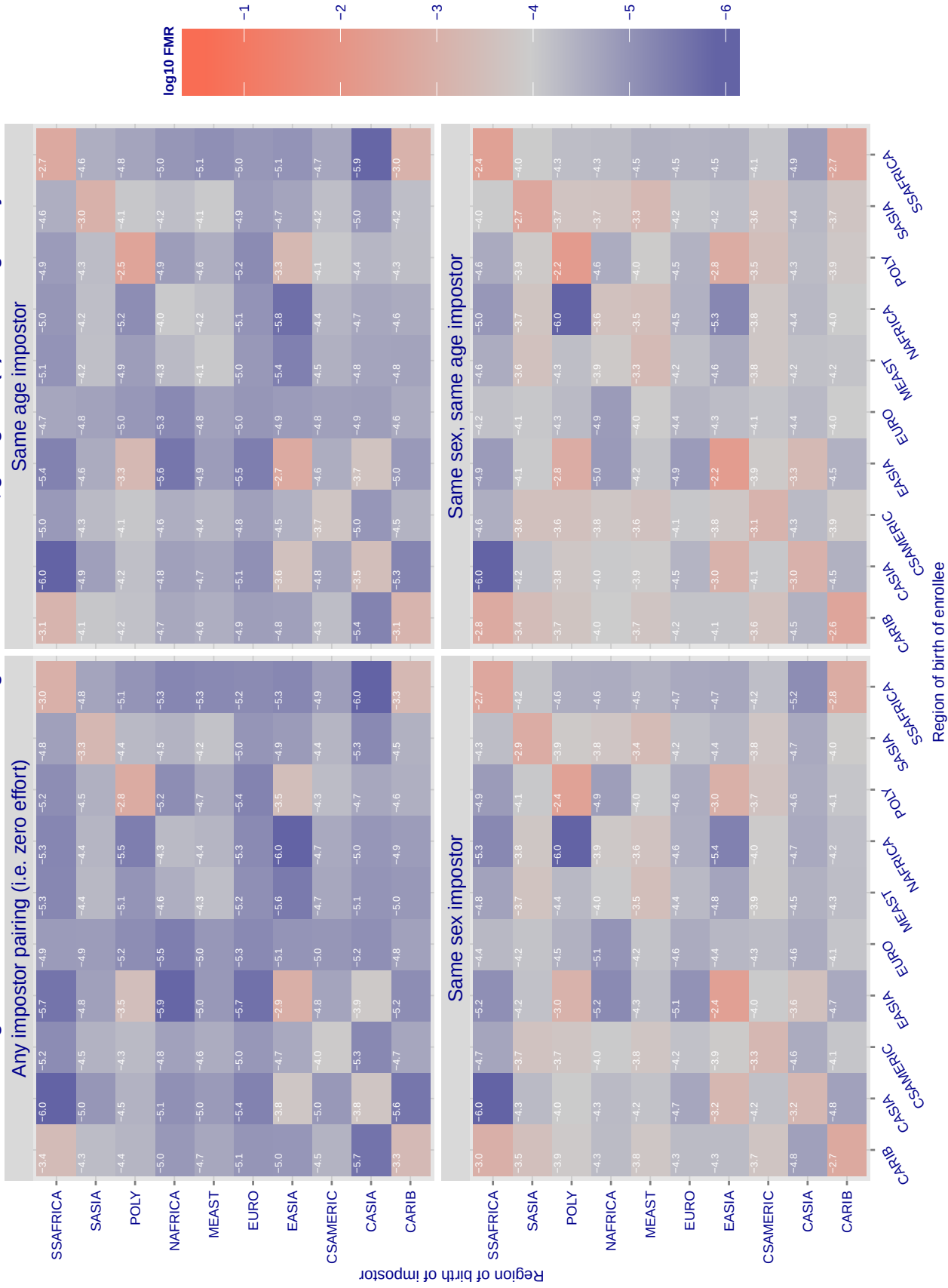
Cross region FMR at threshold $T = 0.867$ for algorithm vocord_002, giving $FMR(T) = 0.0001$ globally.

Figure 34: For algorithm vocord-002 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

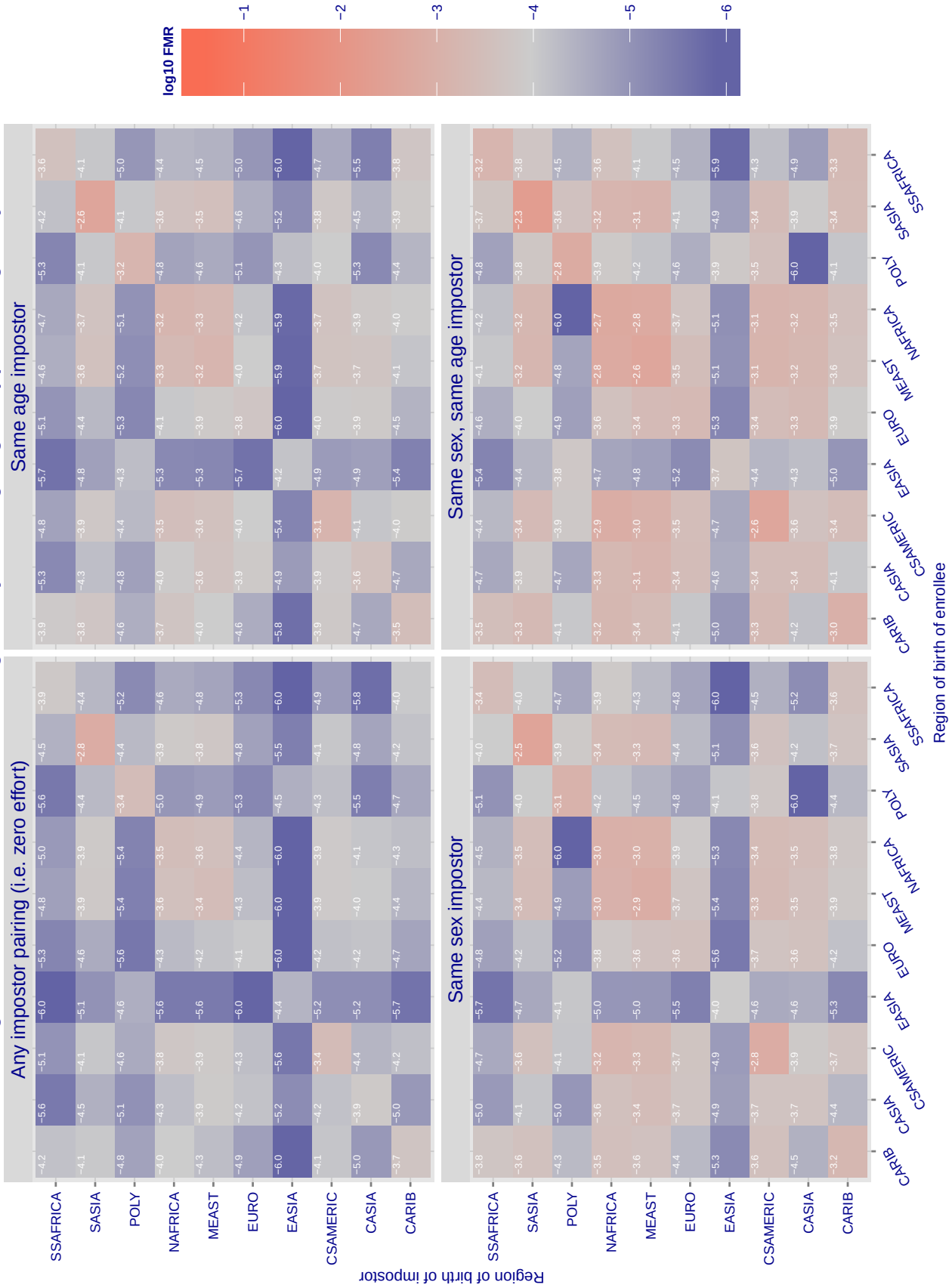
Cross region FMR at threshold $T = 10.098$ for algorithm yitu_000, giving $FMR(T) = 0.0001$ globally.

Figure 35: For algorithm yitu-000 operating on visa images, the heatmap shows false match rates 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 the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 2.869$ for algorithm 3divi_000, giving $FMR(T) = 0.001$ globally.

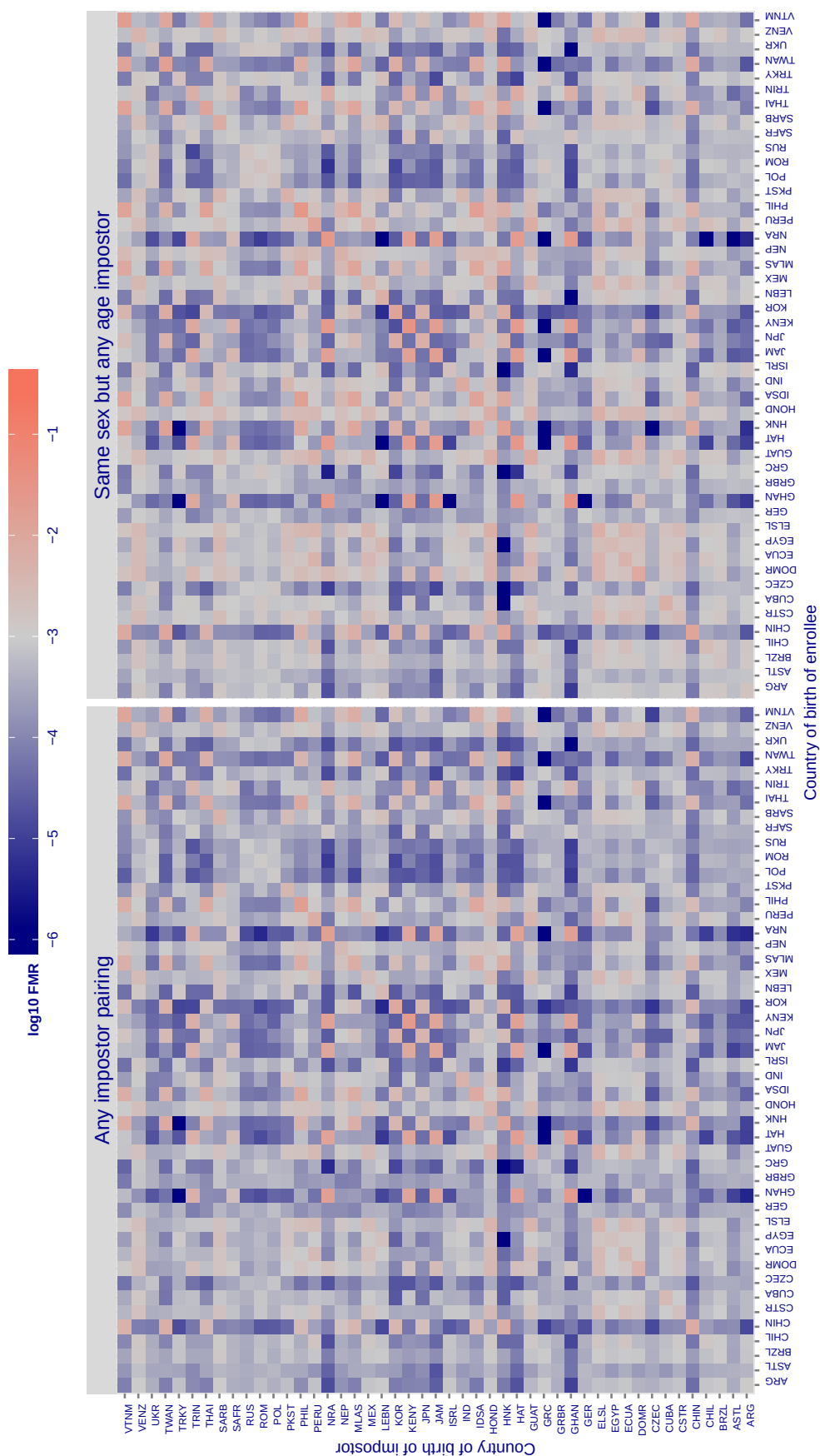


Figure 36: For algorithm 3divi-000 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 78.021$ for algorithm dermalog_001, giving $FMR(T) = 0.001$ globally.

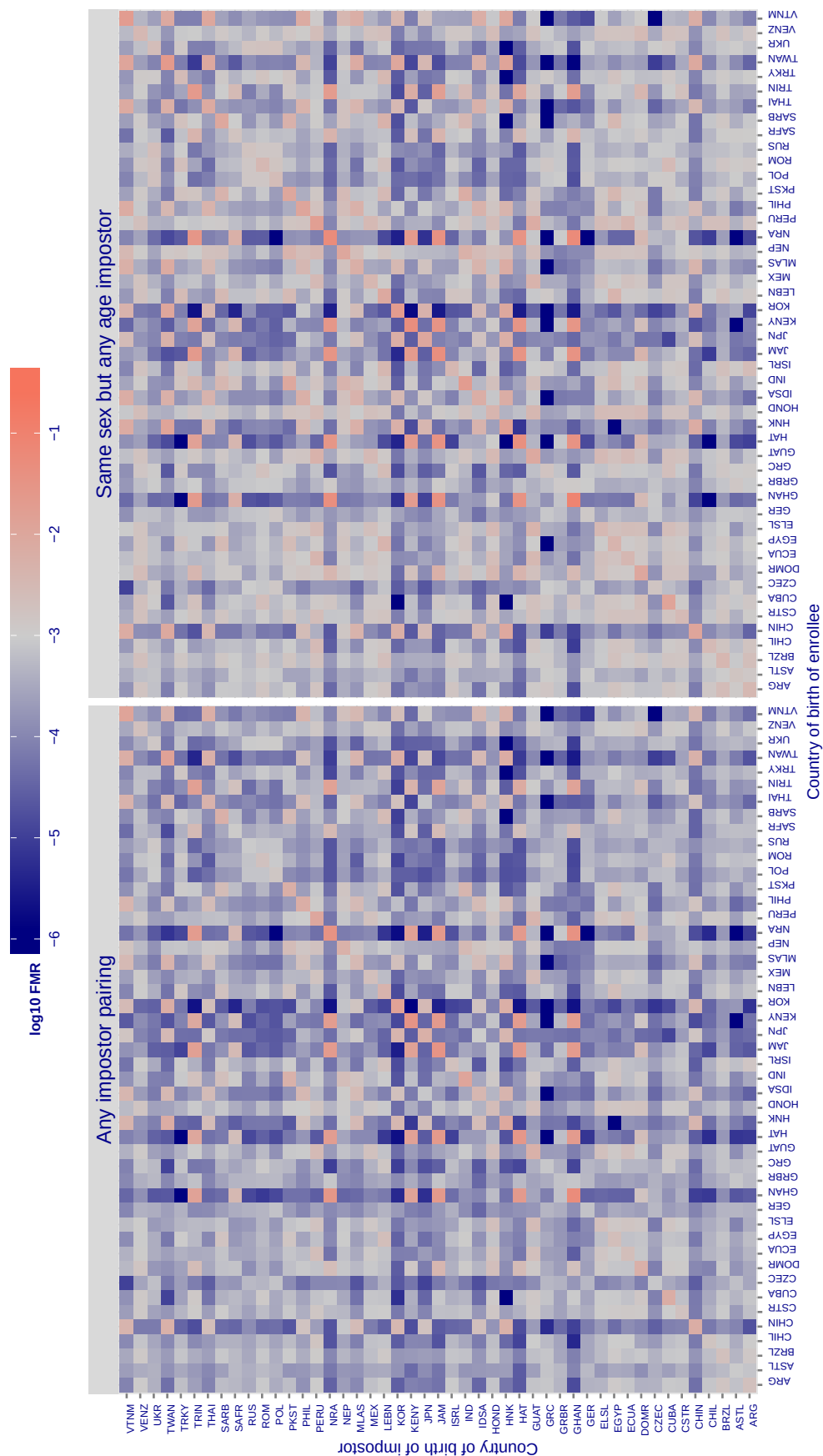


Figure 37: For algorithm dermalog-001 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 78.171$ for algorithm dermalog_002, giving $FMR(T) = 0.001$ globally.

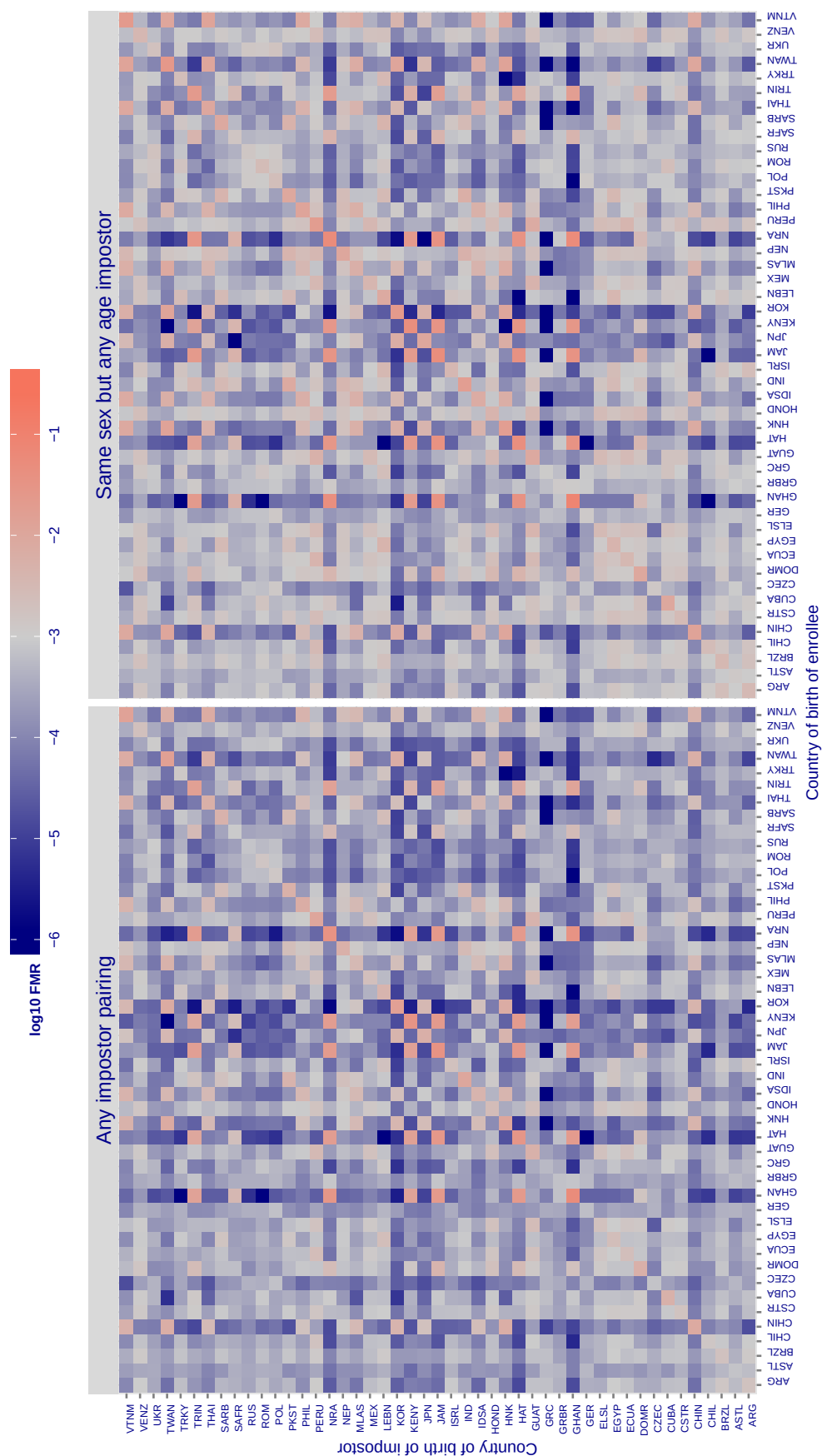


Figure 38: For algorithm dermalog-002 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 30.260$ for algorithm neurotechnology_000, giving $FMR(T) = 0.001$ globally.

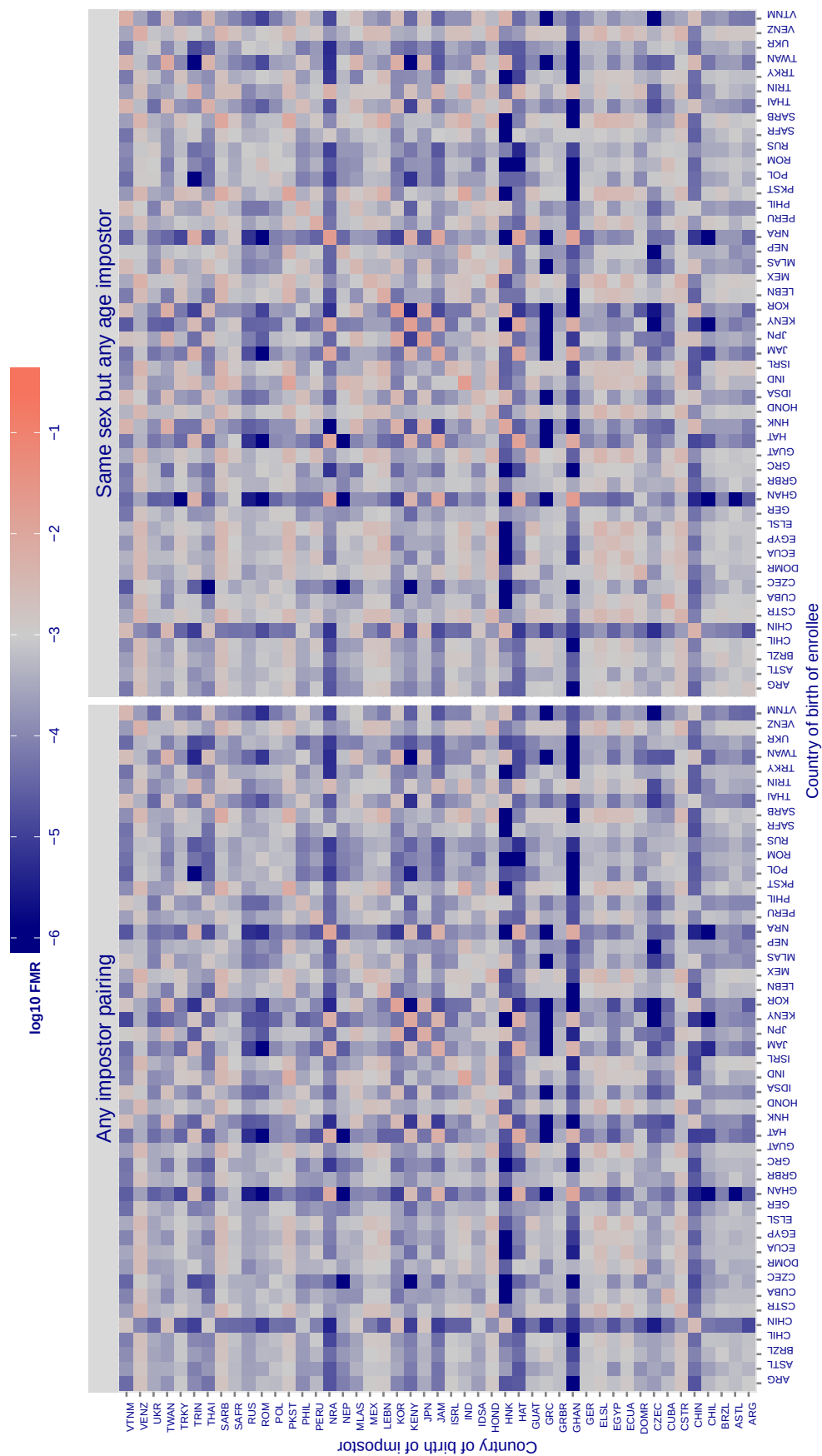


Figure 39: For algorithm neurotechnology-000 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each $+1$ increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 0.091$ for algorithm ntechlab_000, giving $FMR(T) = 0.001$ globally.

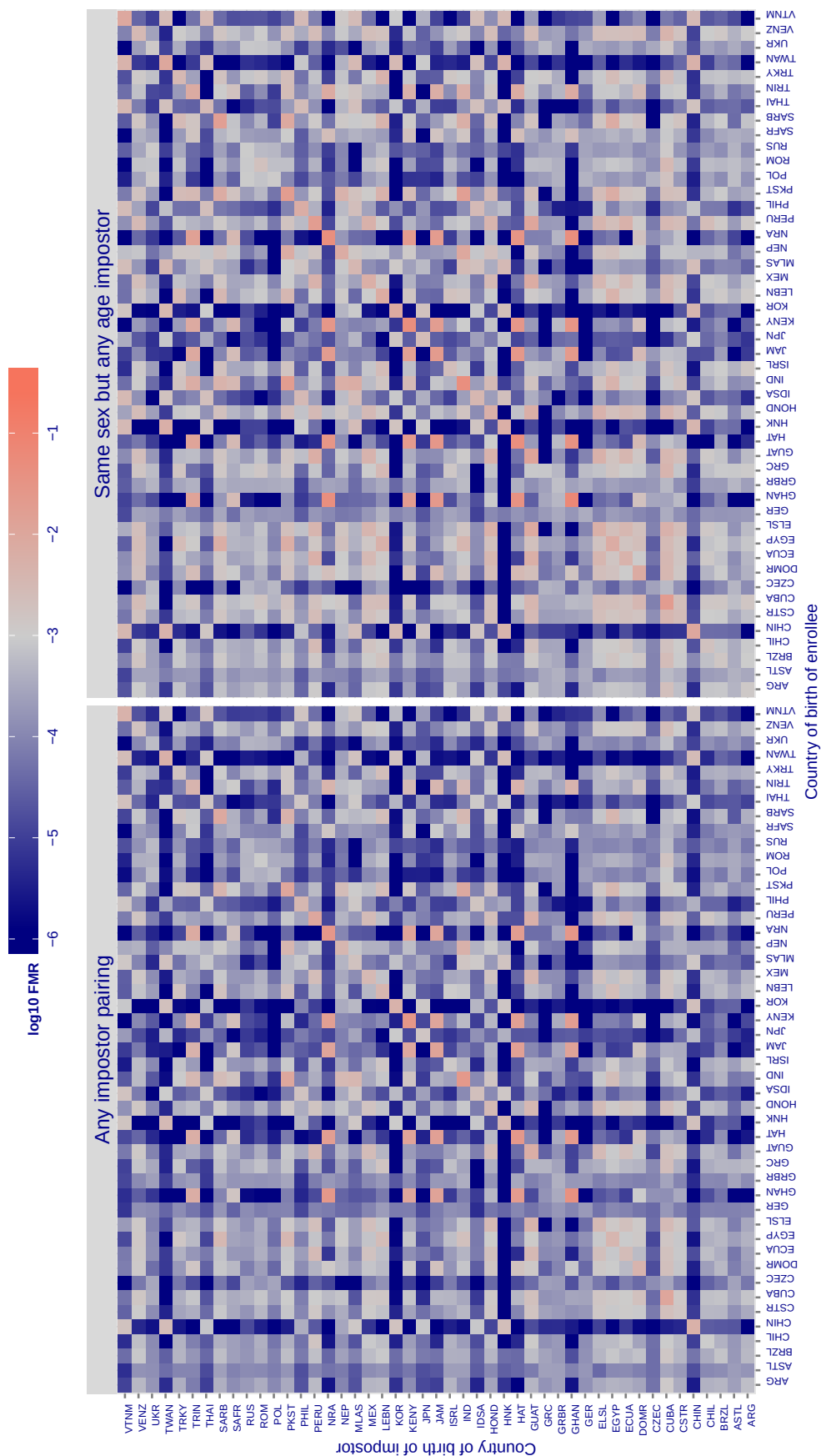


Figure 40: For algorithm ntechlab-000 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 0.089$ for algorithm ntechlab_001, giving $FMR(T) = 0.001$ globally.

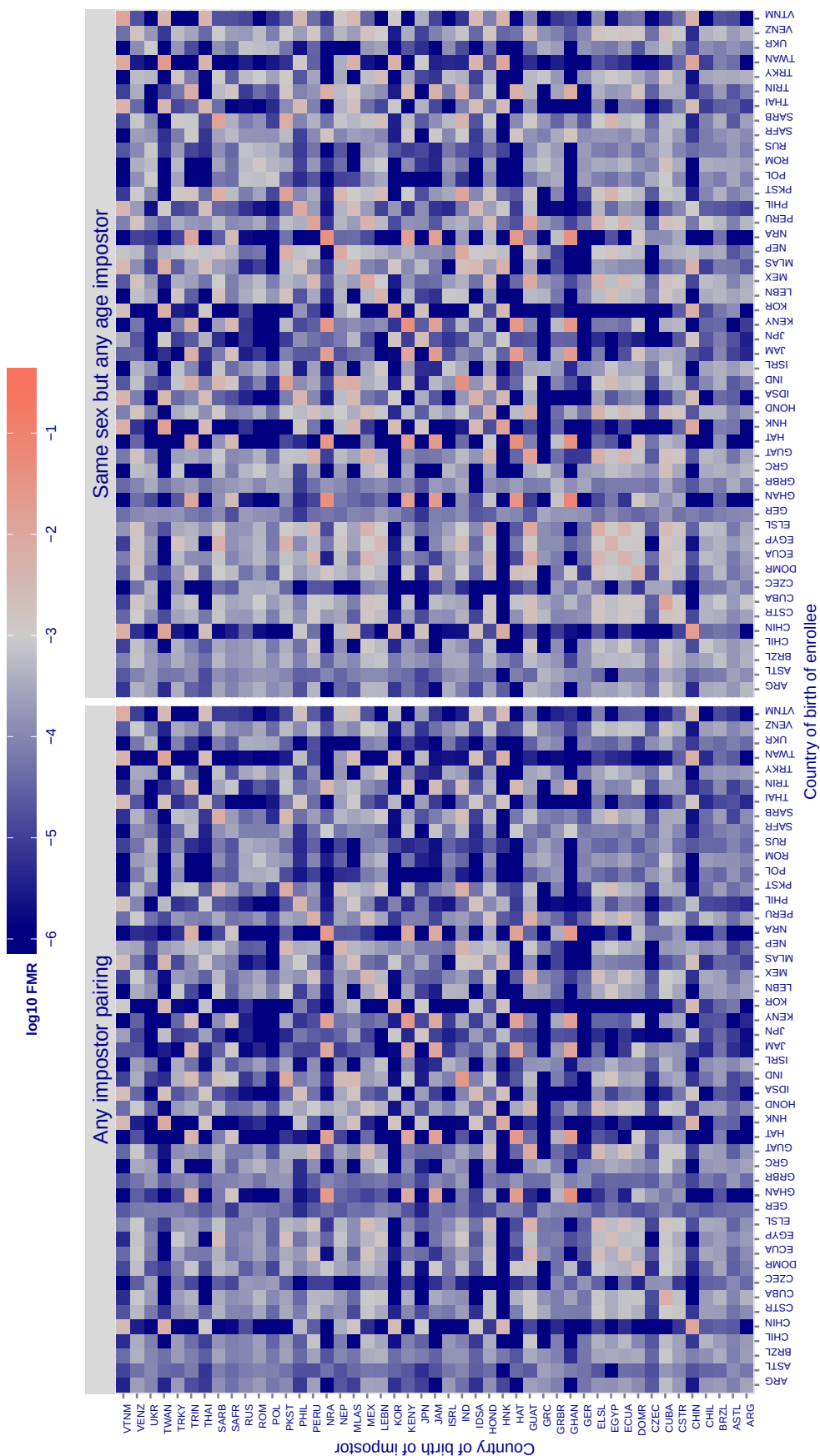


Figure 41: For algorithm ntechlab-001 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 0.582$ for algorithm rankone_000 , giving $\text{FMR}(T) = 0.001$ globally.

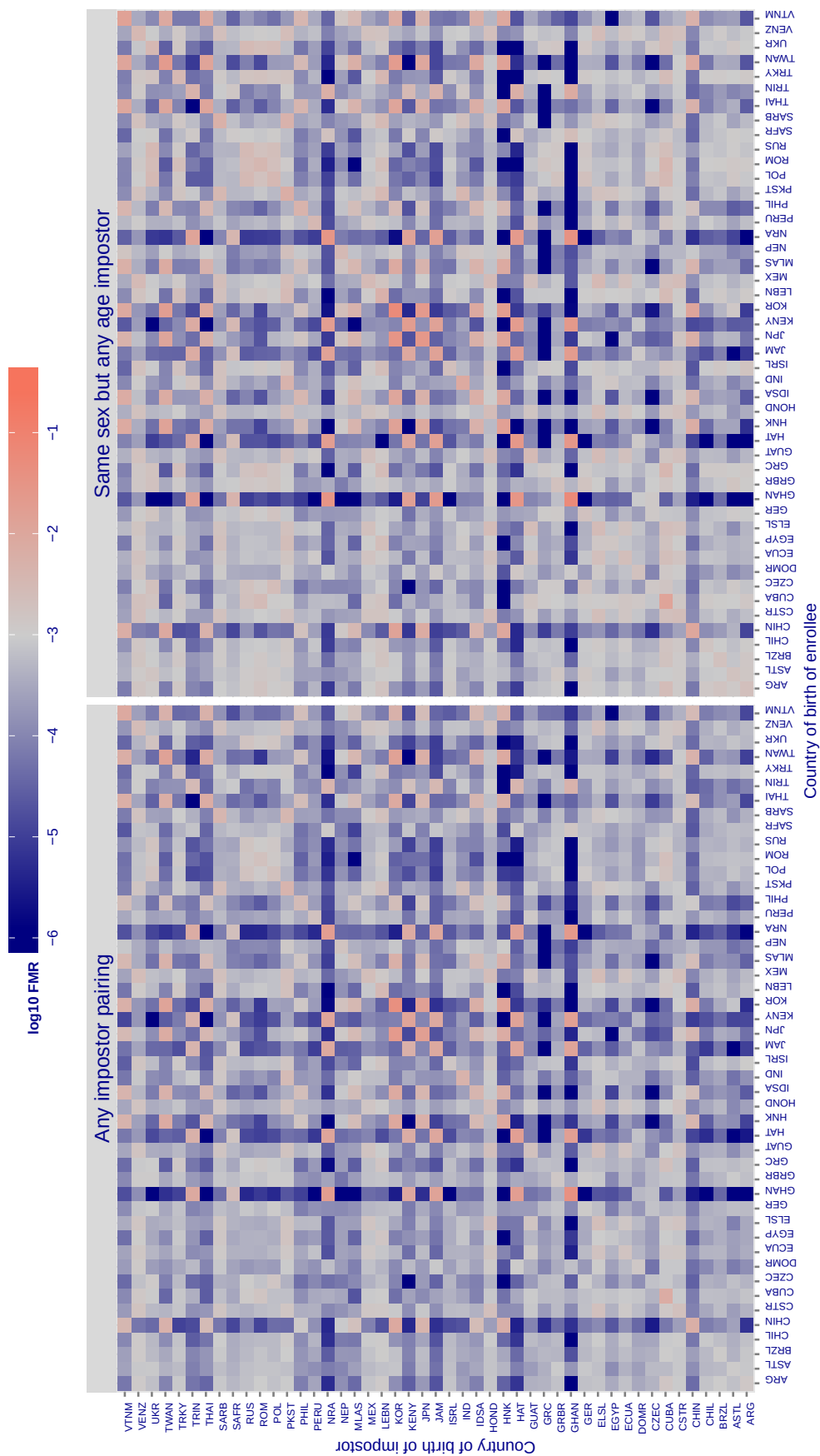


Figure 42: For algorithm rankone_000 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each $+1$ increase in $\log_{10}$ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 0.614$ for algorithm rankone_001, giving $FMR(T) = 0.001$ globally.

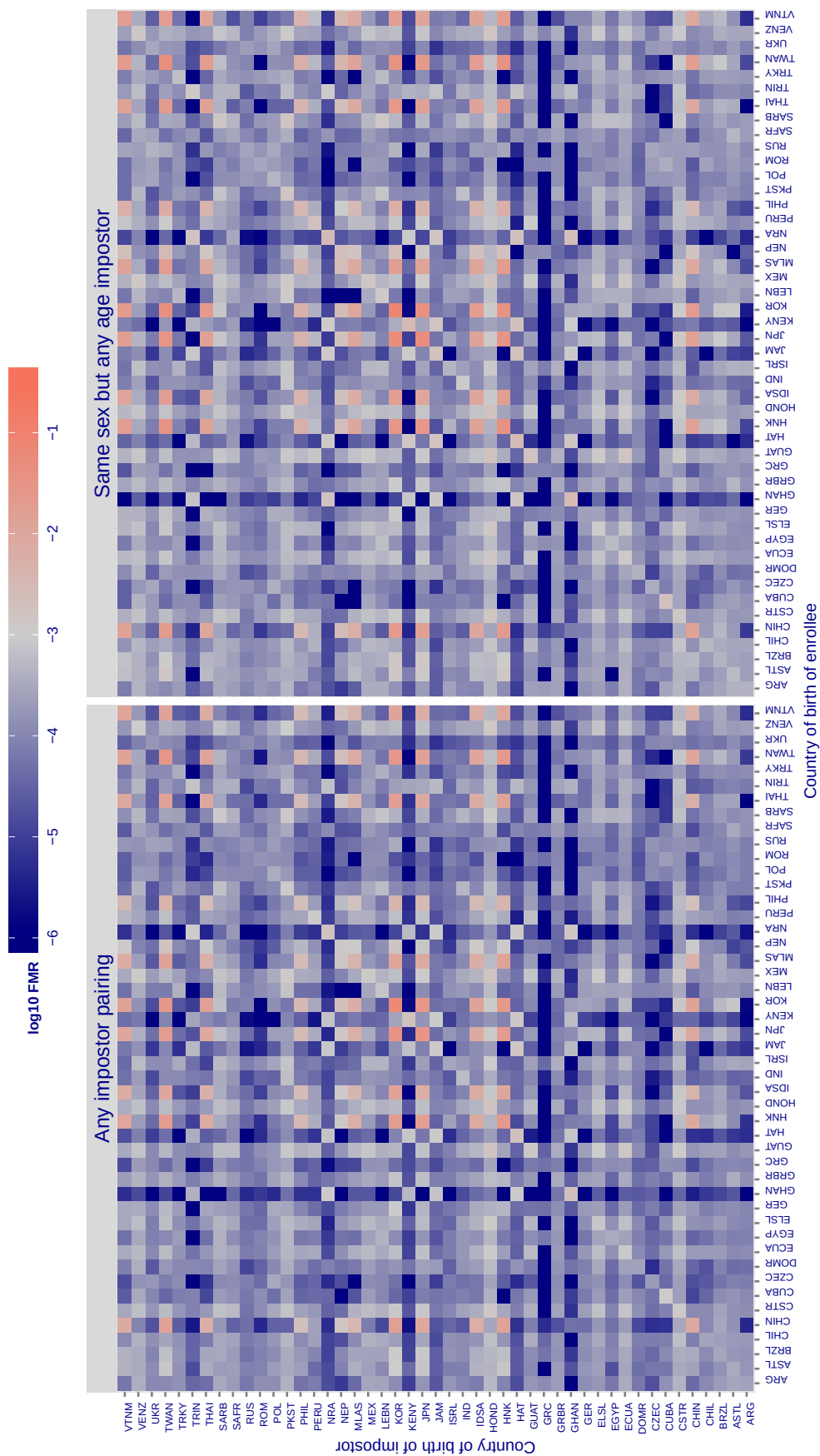


Figure 43: For algorithm rankone-001 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 74.060$ for algorithm `samtech_000`, giving $FMR(T) = 0.001$ globally.

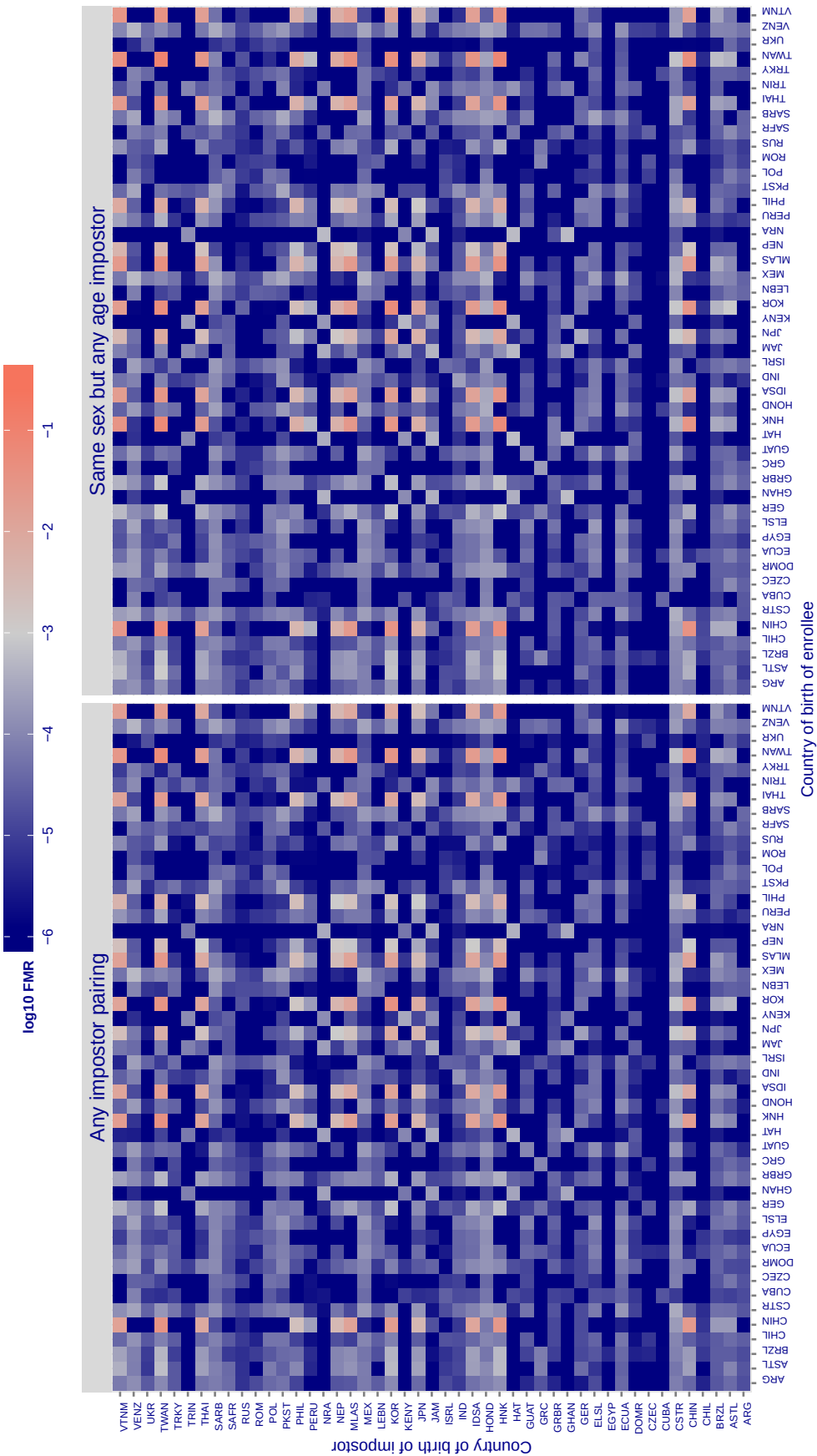


Figure 44: For algorithm `samtech-000` operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each $+1$ increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 9.972$ for algorithm tongyitrans_001, giving $FMR(T) = 0.001$ globally.

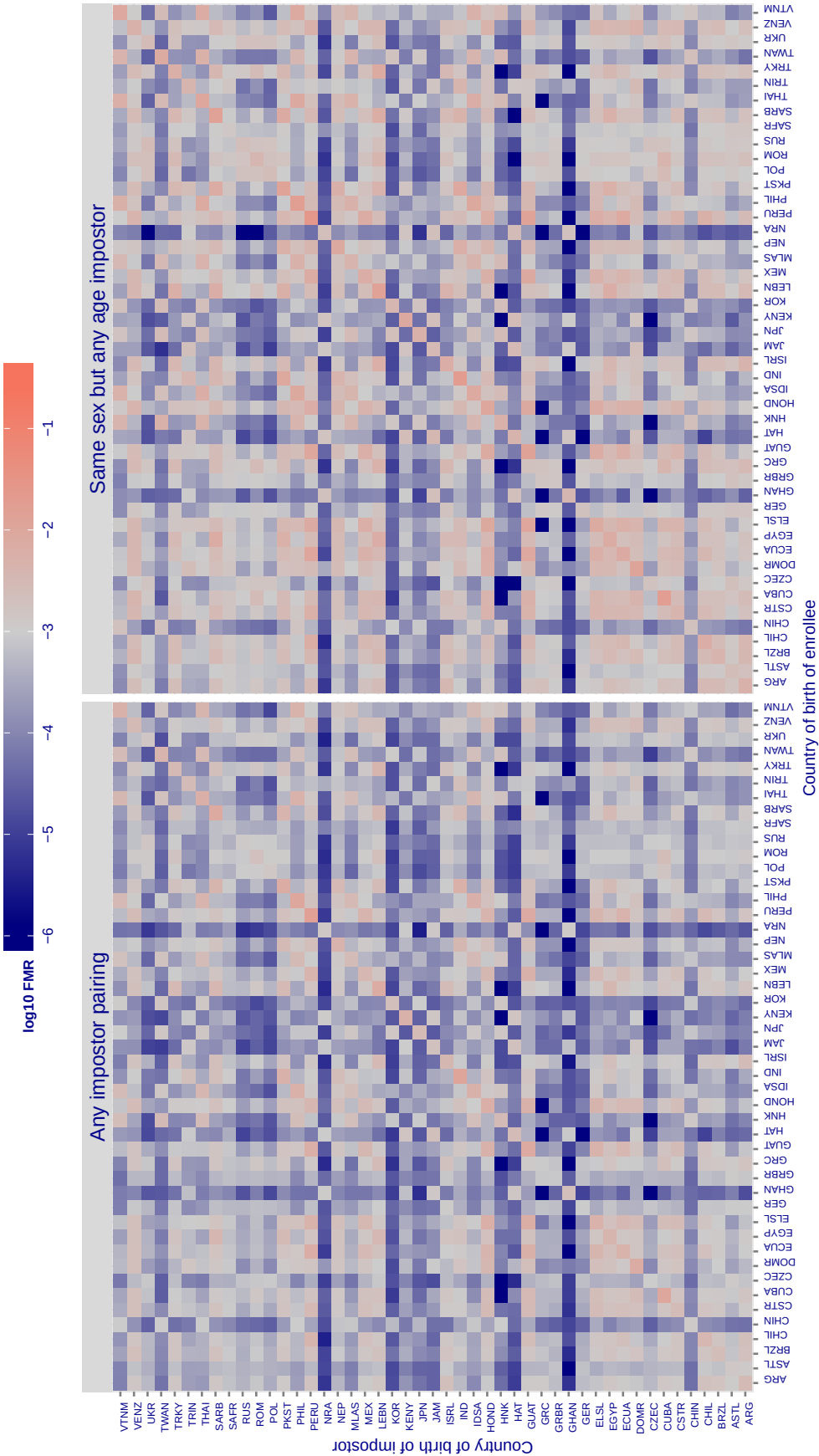


Figure 45: For algorithm tongyitrans-001 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in log₁₀ FMR corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 1.000$ for algorithm tuple_001, giving $FMR(T) = 0.001$ globally.

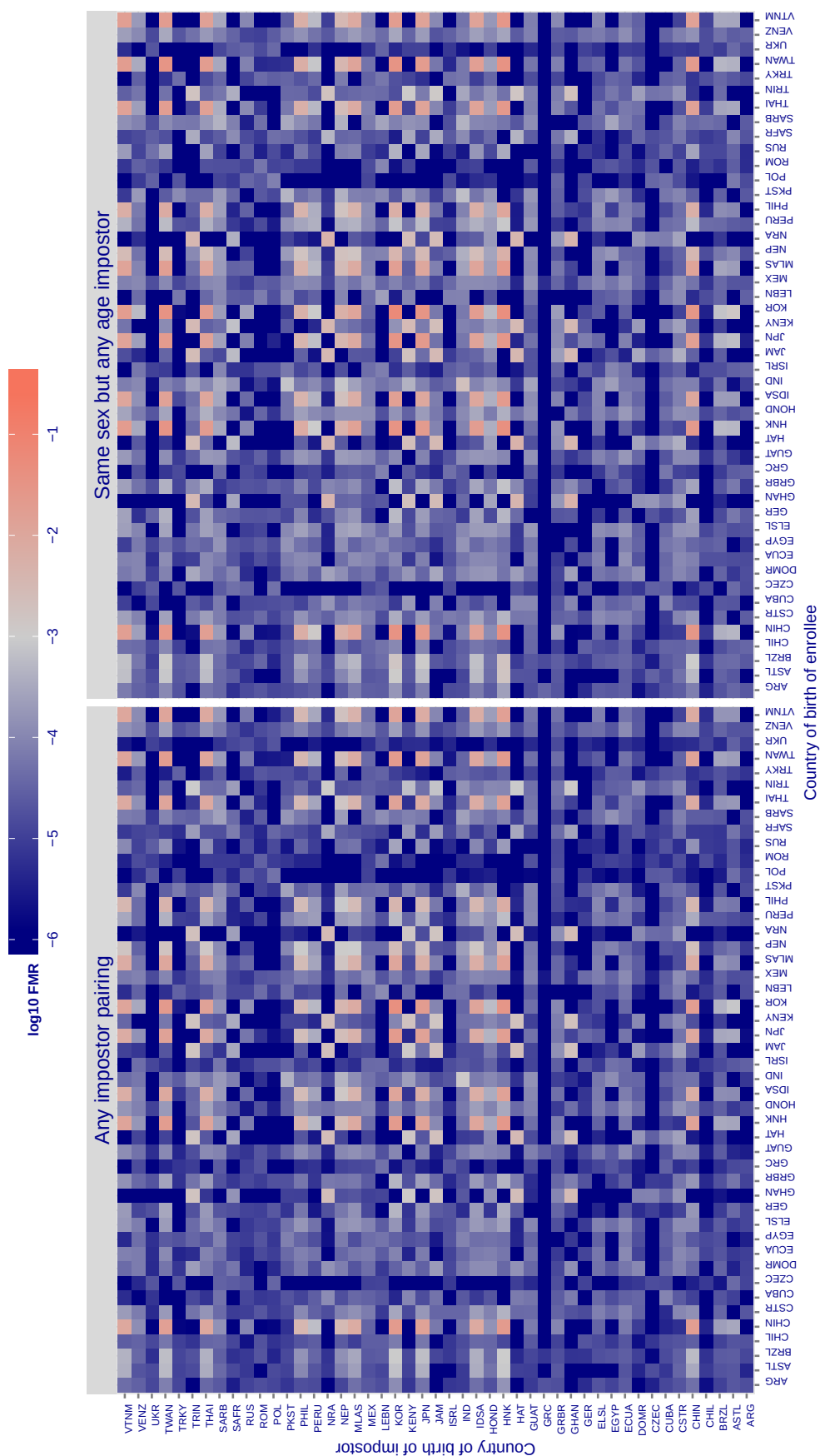


Figure 46: For algorithm tuple-001 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 16.410$ for algorithm vcog_001, giving $FMR(T) = 0.001$ globally.

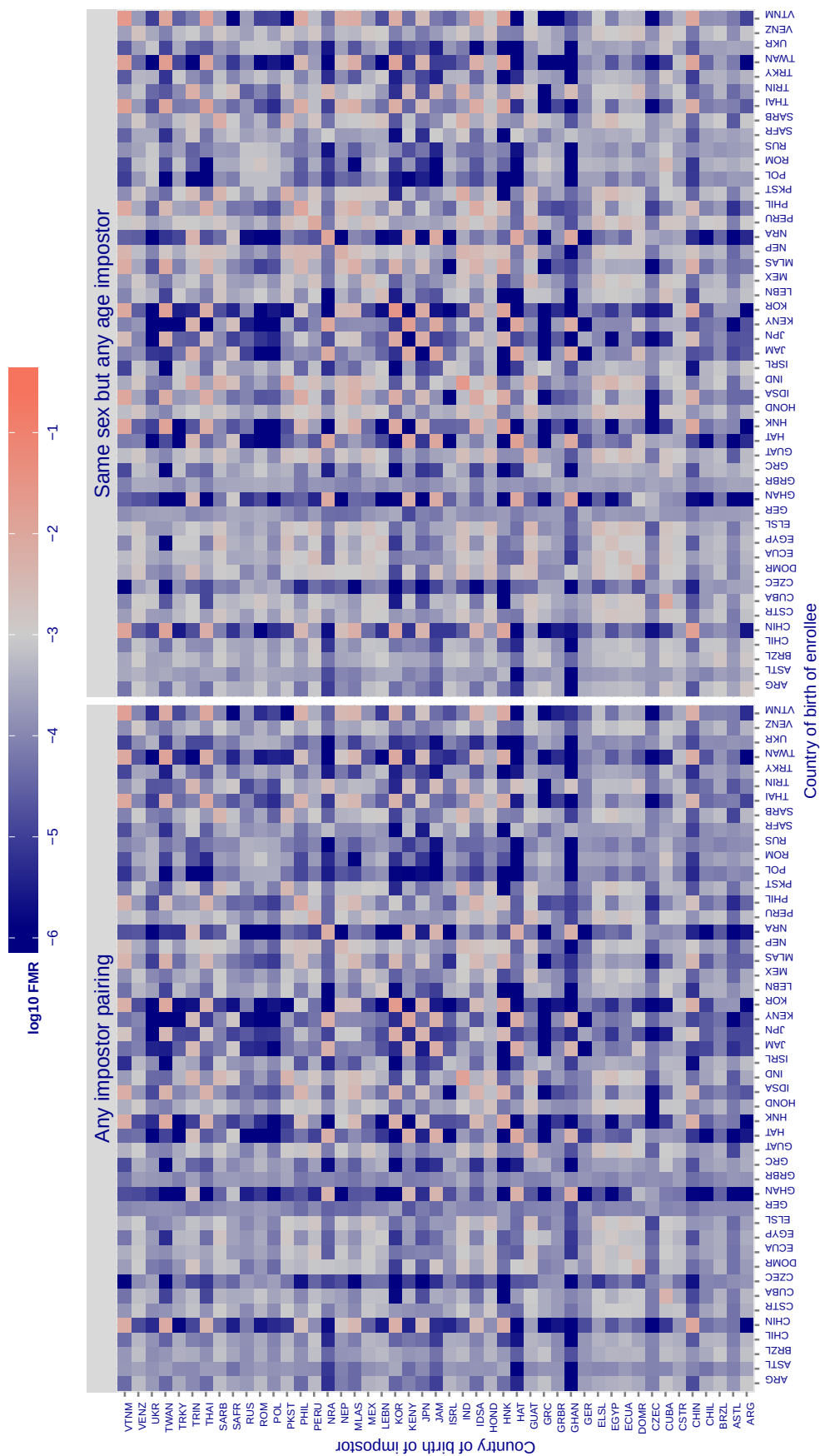


Figure 47: For algorithm vcog-001 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

Cross country FMR at threshold $T = 0.095$ for algorithm `vigilantsolutions_000`, giving $FMR(T) = 0.001$ globally.

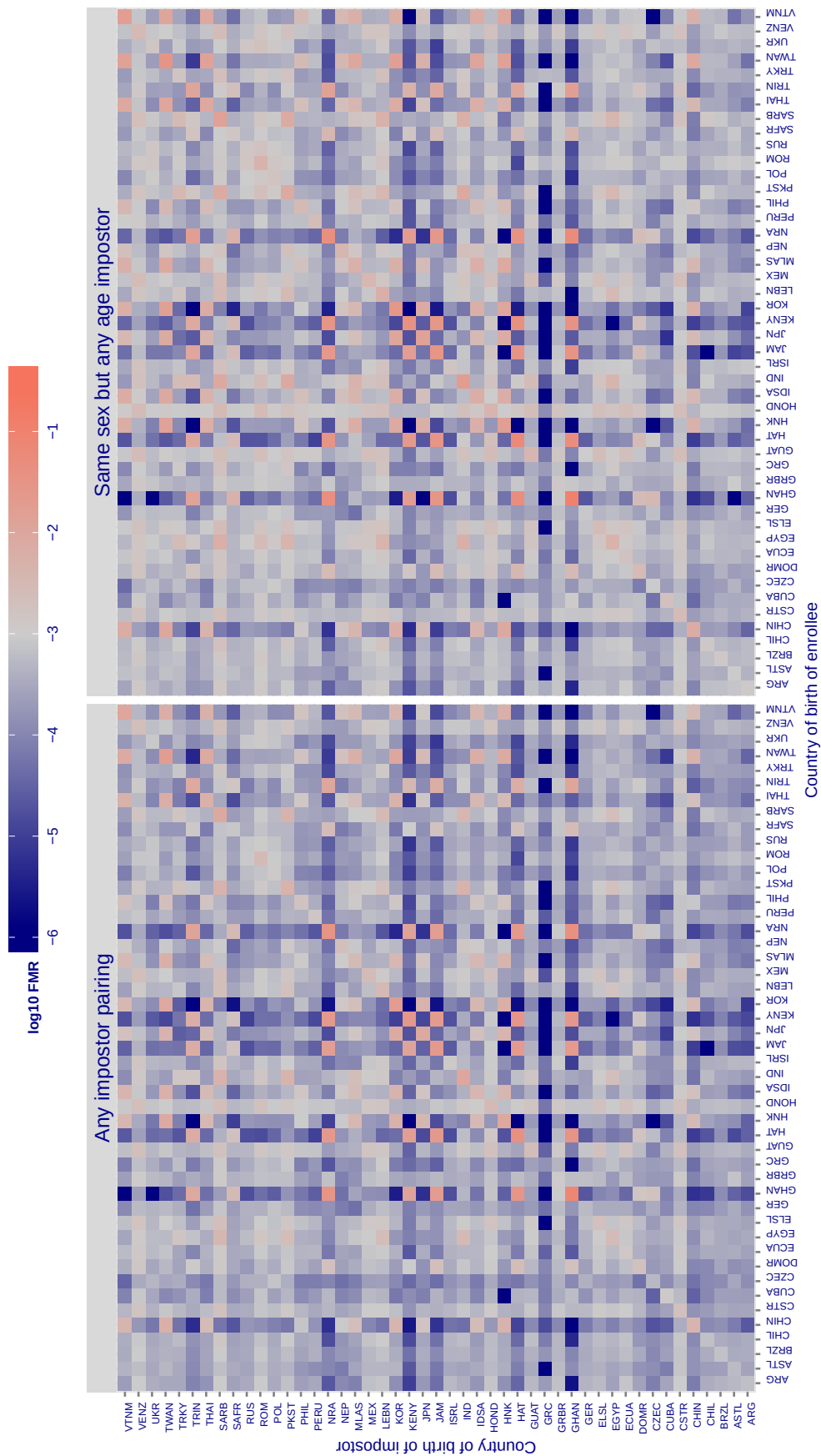


Figure 48: For algorithm `vigilantsolutions-000` operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each $+1$ increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

The figure displays two heatmaps side-by-side, comparing the log10 False Match Rate (FMR) for different impostor pairing scenarios across 48 countries. The color scale ranges from -6 (dark blue) to -1 (dark red), with intermediate values represented by lighter shades of blue and green.

Left Heatmap: Any impostor pairing
 This heatmap shows the log10 FMR for any impostor pairing. The y-axis is labeled "Country of birth of impostor" and the x-axis is labeled "Country of birth of enrollee". The diagonal elements, representing self-pairings, are consistently dark blue, indicating the lowest FMR values (around -6). The off-diagonal elements show varying levels of FMR, with some countries exhibiting higher rates (lighter colors) than others.

Right Heatmap: Same sex but any age impostor
 This heatmap shows the log10 FMR for impostor pairing where the impostor is of the same sex as the enrollee but of any age. The overall pattern is similar to the left heatmap, with dark blue diagonals and varying off-diagonal values. The FMR values generally appear slightly higher (lighter colors) compared to the "Any impostor pairing" scenario.

Country List (Y-axis for both):
 VTNM, VENZ, UKR, TWAN, TRKY, TRIN, THAI, SARB, SAFR, RUS, ROM, POL, PKST, PHIL, PERU, NRA, NEP, MEX, MEX, LEBN, KOR, KENY, JPN, JAM, ISRL, IND, IDSA, HOND, HNK, HAT, GUAT, GRC, GRBR, GHAN, GER, ELSL, EGY, ECUA, DOMR, CZEC, CUBA, CSTR, CHIN, CHIL, BRZL, ARG.

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

Cross country FMR at threshold $T = 9.942$ for algorithm yitu_000, giving $FMR(T) = 0.001$ globally.

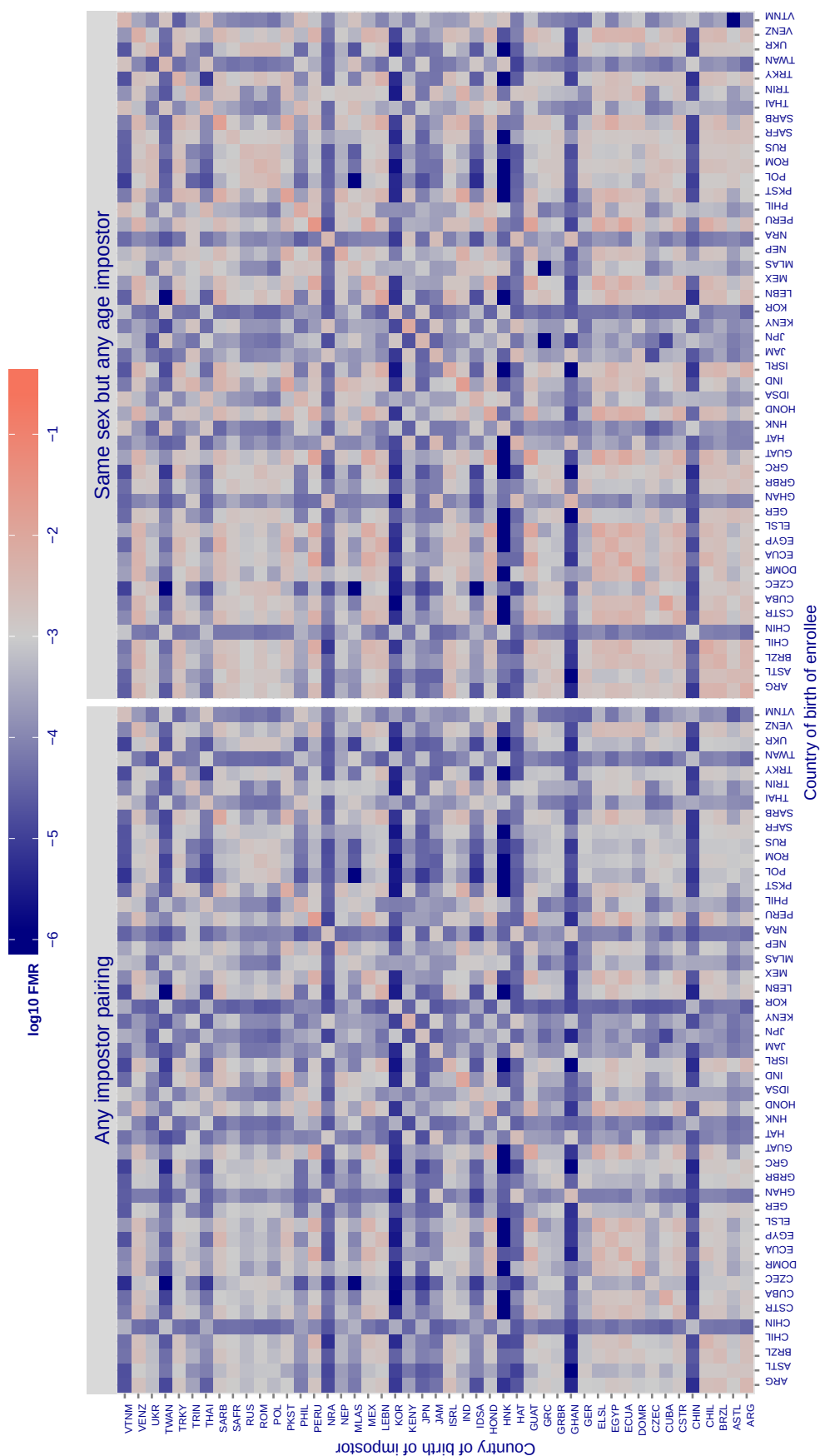


Figure 50: For algorithm yitu-000 operating on visa images, the heatmap shows false match rates observed over impostor comparisons of faces from different individuals who were born in the given country pair. False matches are counted against a recognition threshold fixed globally to give the target FMR in the plot title, computed over all $O(10^{10})$ impostor comparisons. If text appears in each box it give the same quantity as that coded by the color. Grey indicates FMR is at the intended FMR target level. Light red colors present a security vulnerability to, for example, a passport gate. Each +1 increase in $\log_{10} FMR$ corresponds to a factor of 10 increase in FMR. The matrix is not quite symmetric because images in the enrollment and verification sets are different.

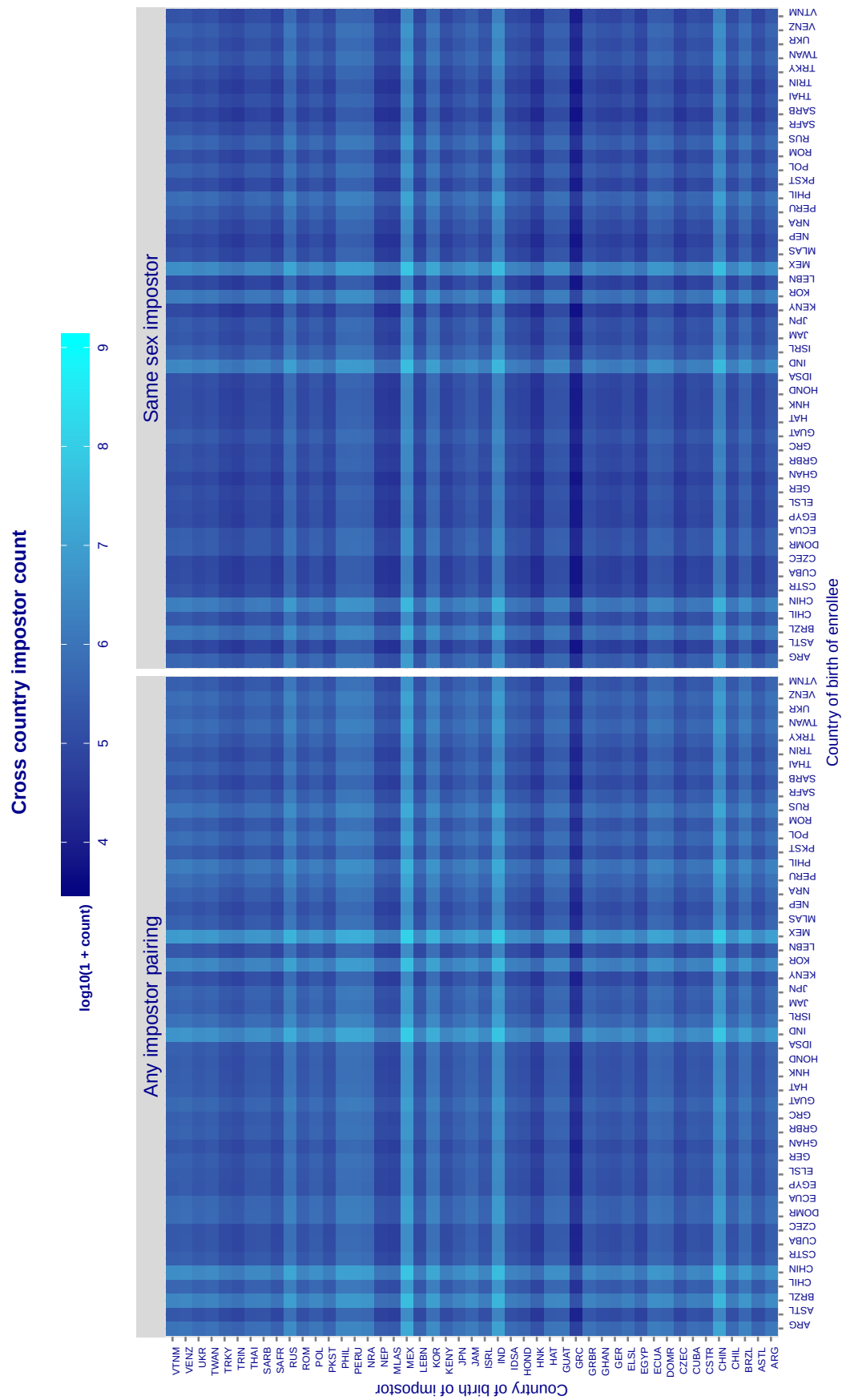


Figure 51: For visa images, the heatmap shows The count of impostor comparisons of faces from different individuals who were born in the given country pair.

4.6.2 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 and same region impostors are more successful. On the diagonal, an impostor is more likely to succeed by posing as someone of the same sex. If $\Delta \log_{10} \text{FMR} = 0.2$, then same-sex same-region FMR exceeds the all-pairs FMR by factor of $10^{0.2} = 1.6$.
- ▷ 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 $\Delta \log_{10} \text{FMR} = 1.0$ 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) $\text{FMR} = 0.001$.
- ▷ Algorithms vary.
- ▷ We computed the same quantities for a global $\text{FMR} = 0.0001$. The effects are similar.

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

Cross age FMR at threshold $T = 3.057$ for algorithm 3divi_000, giving $FMR(T) = 0.0001$ globally.

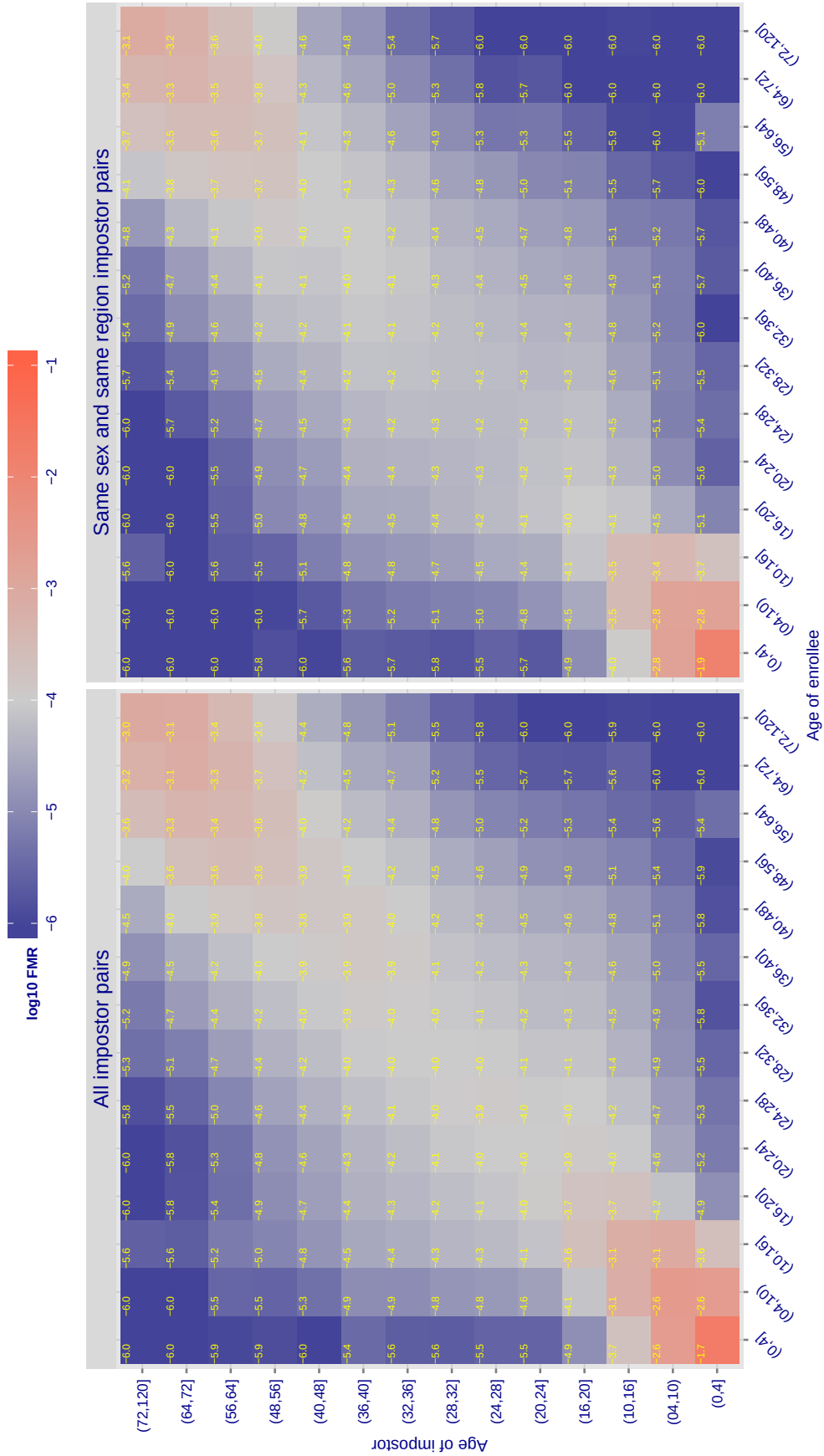


Figure 52: For algorithm 3divi-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 81.064$ for algorithm dermalog_001, giving $FMR(T) = 0.0001$ globally.

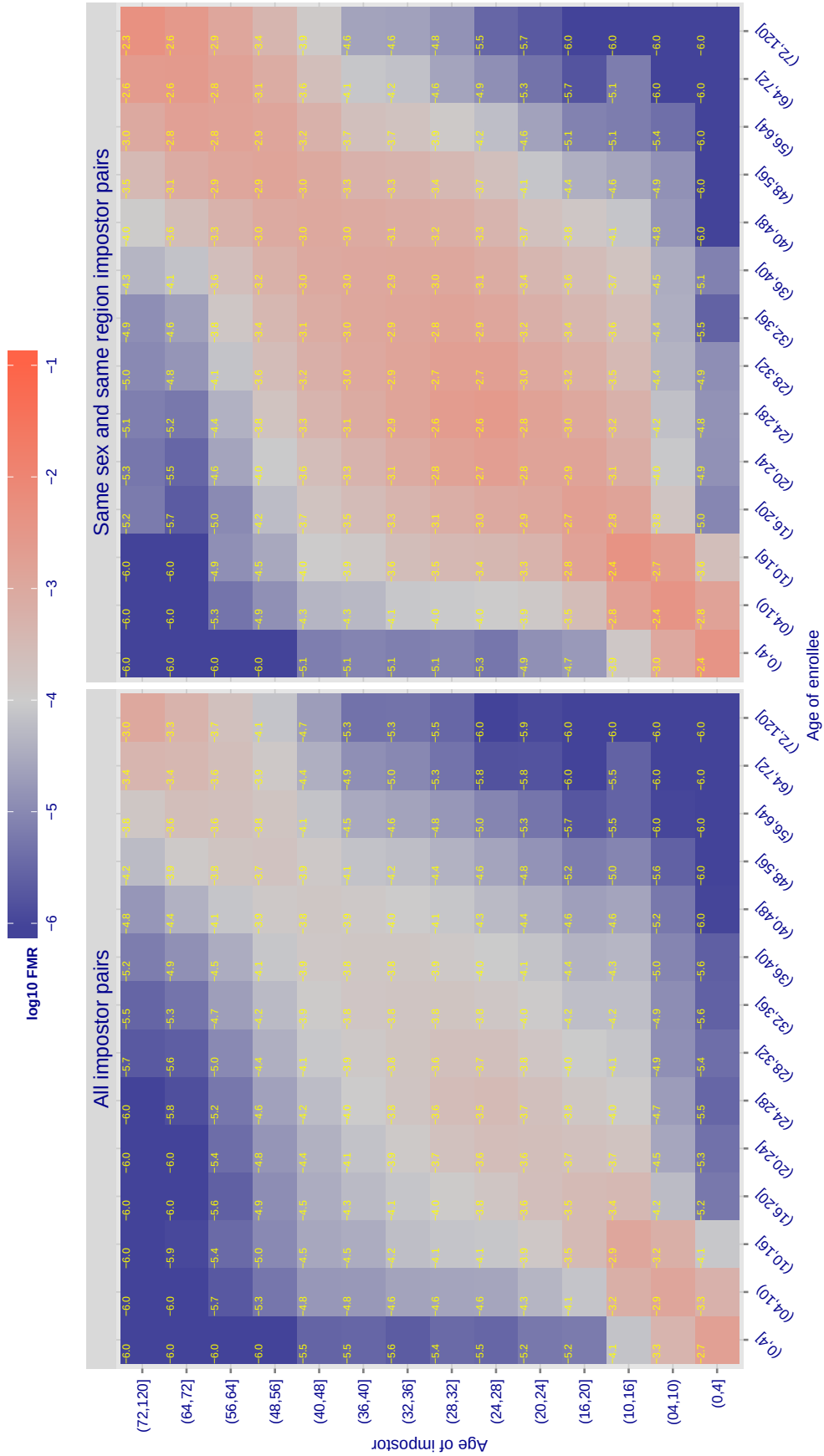


Figure 53: For algorithm dermalog-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 81.164$ for algorithm dermalog_002, giving $FMR(T) = 0.0001$ globally.

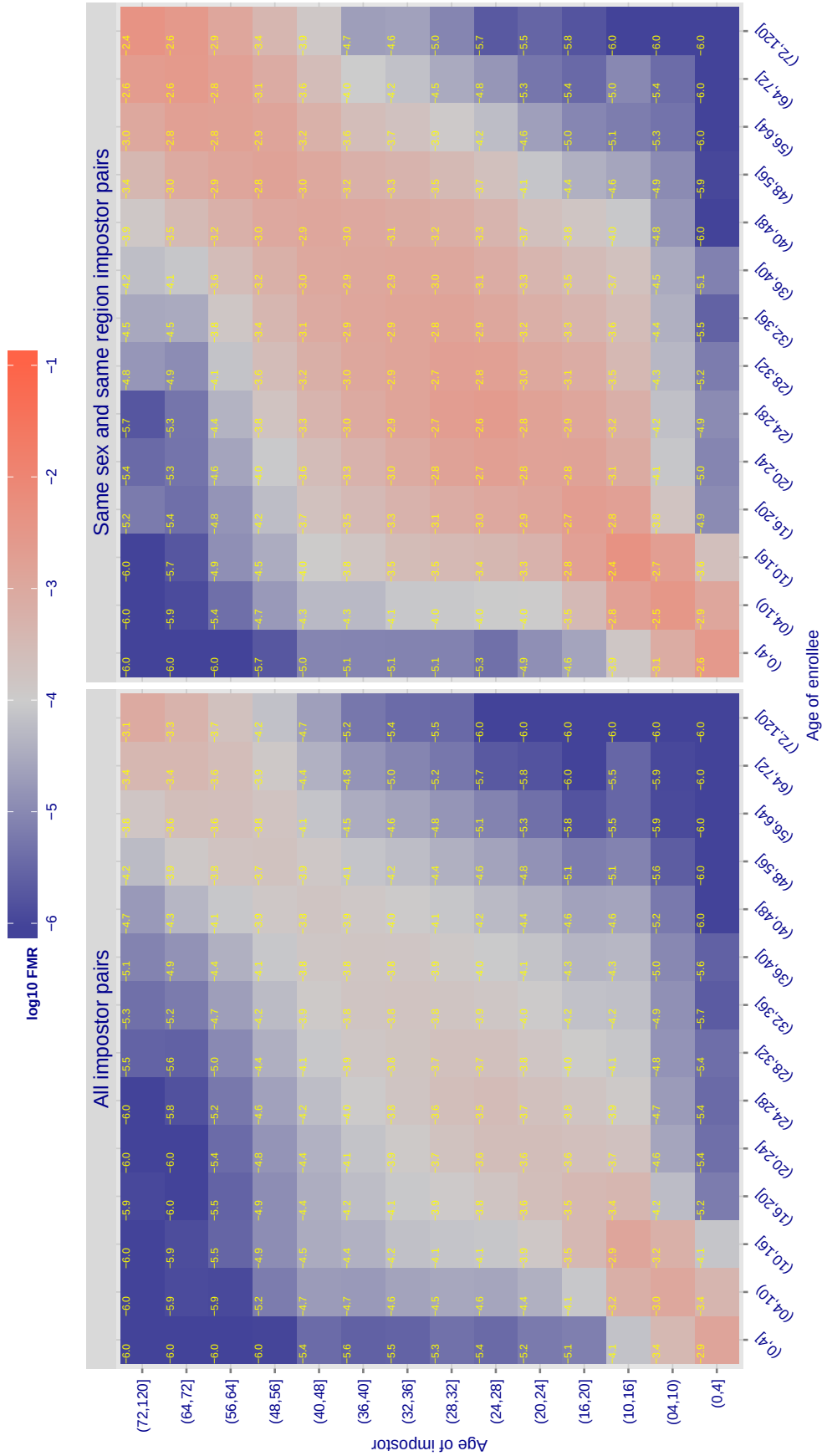


Figure 54: For algorithm dermalog-002 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

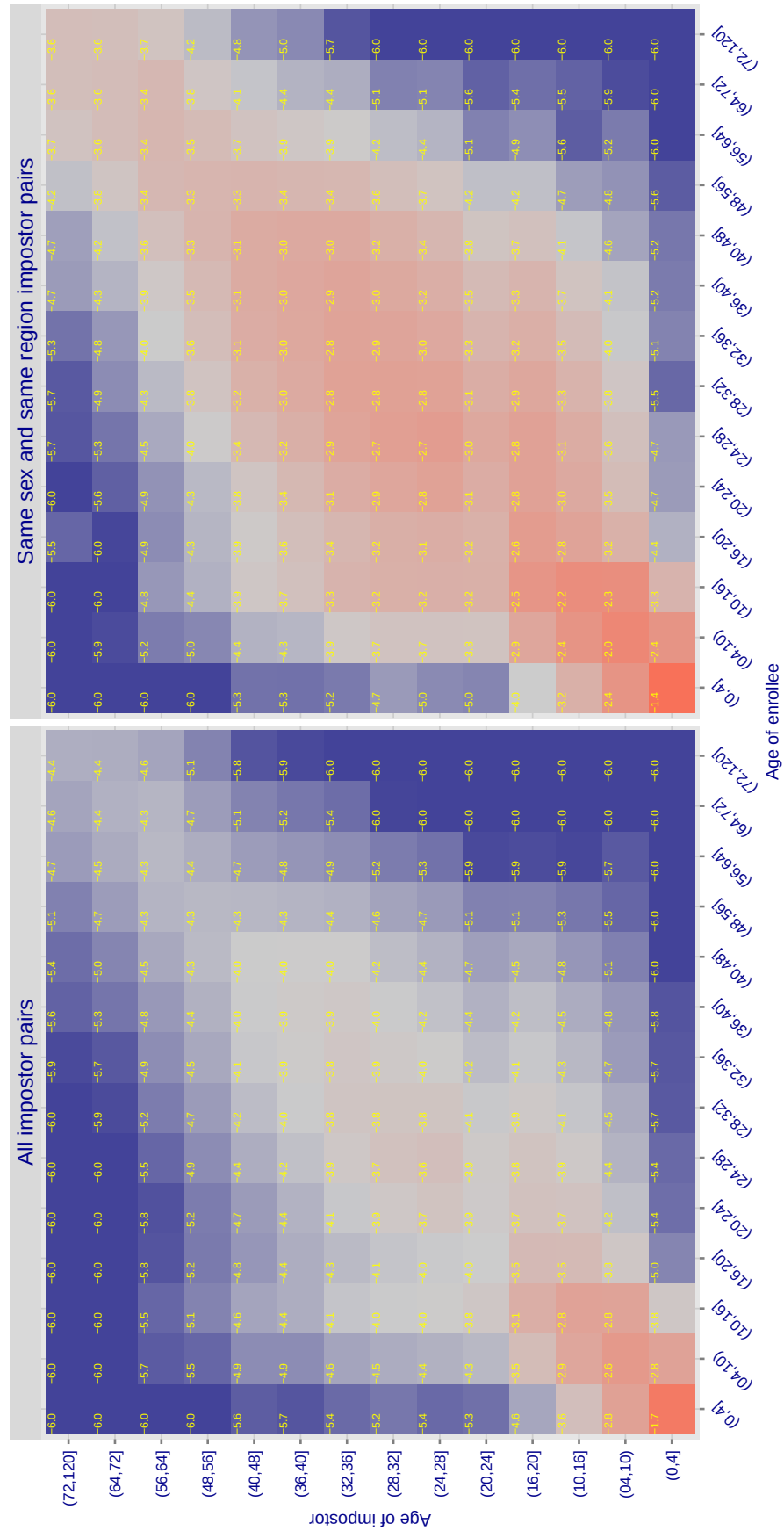


Figure 55: For algorithm digitalbarriers-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 990.194$ for algorithm itmo_001 , giving $\text{FMR}(T) = 0.0001$ globally.

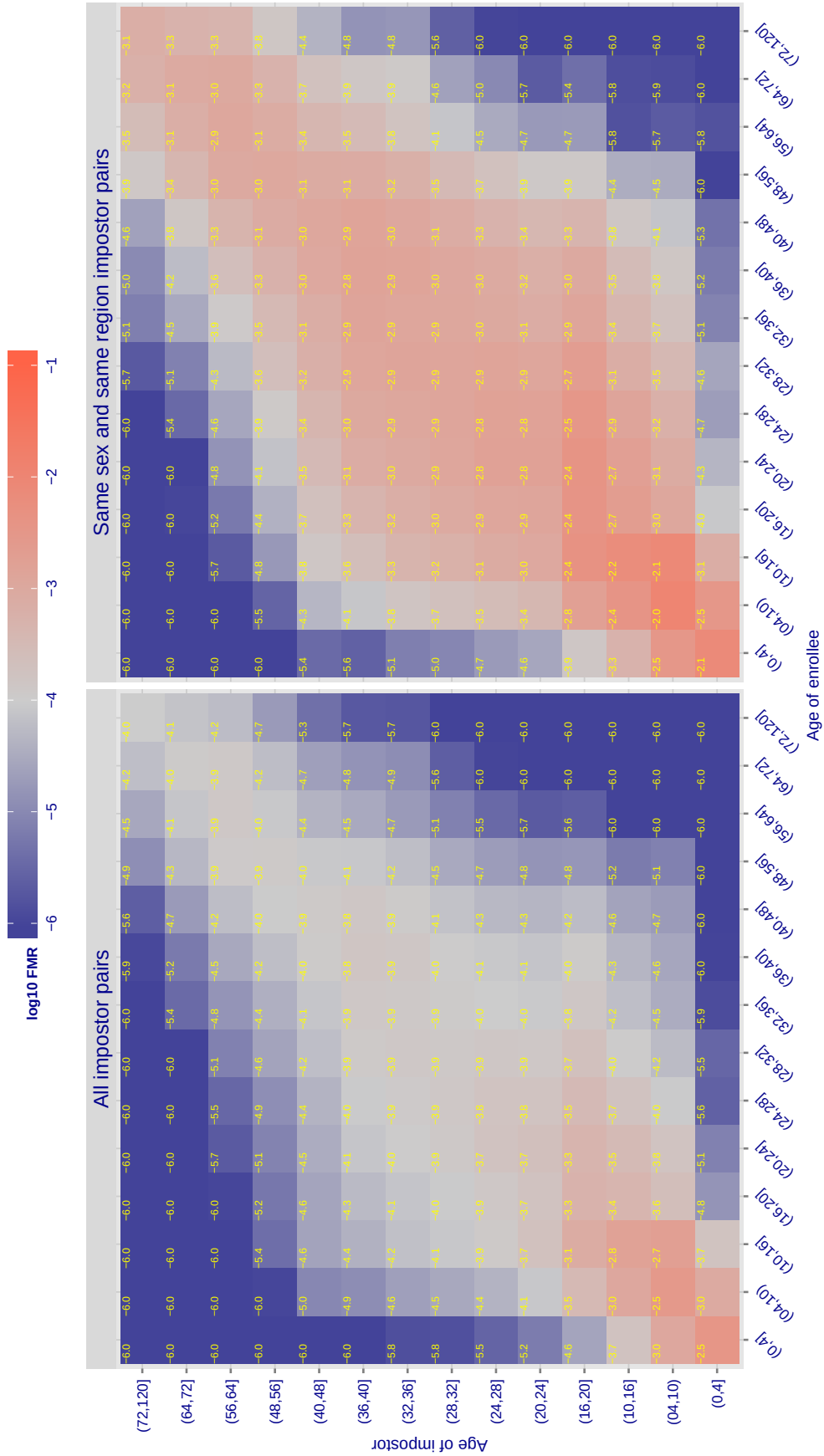


Figure 56: For algorithm itmo_001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $\text{FMR} = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 43.010$ for algorithm neurotechnology_000, giving $FMR(T) = 0.0001$ globally.

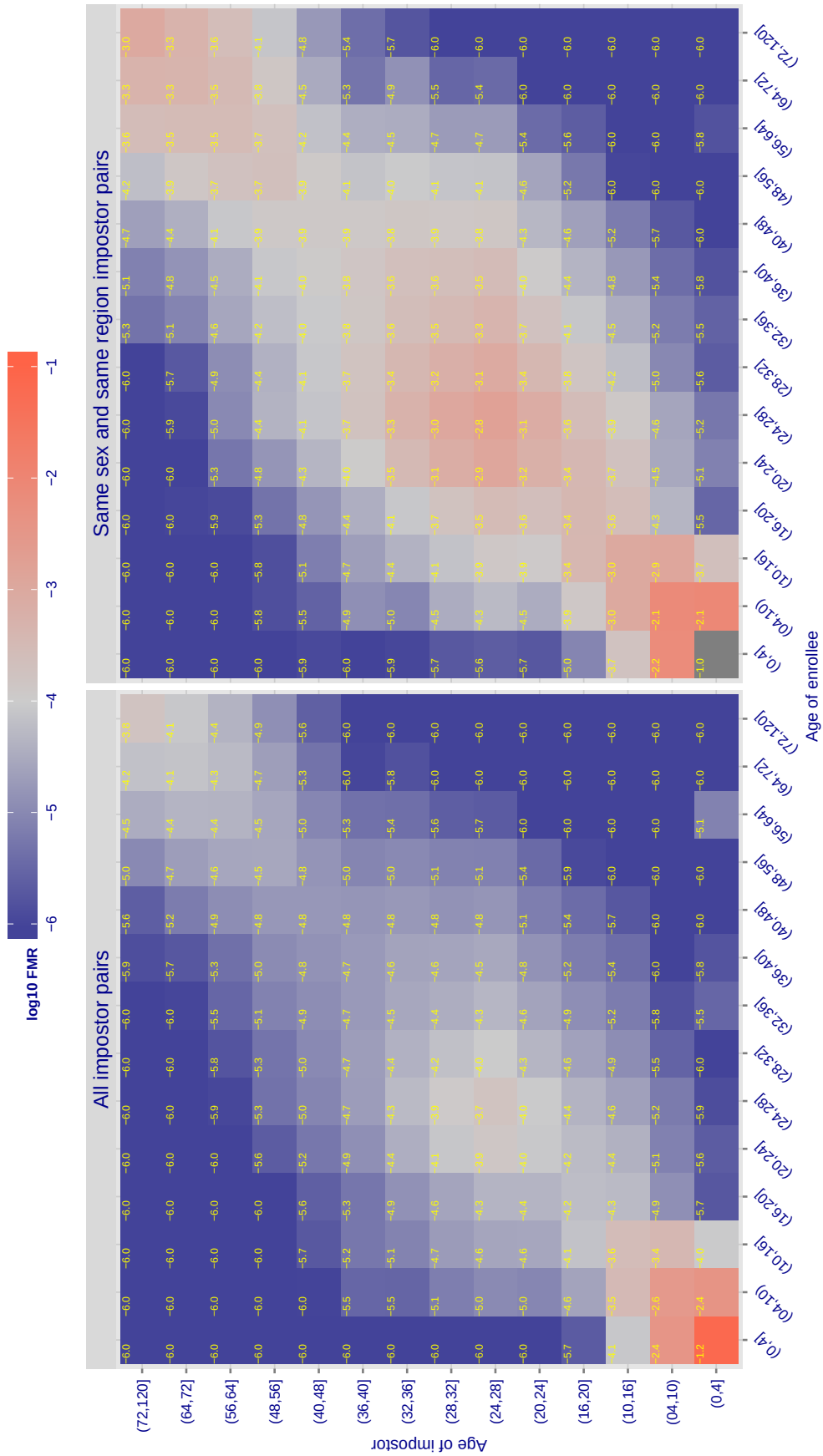


Figure 57: For algorithm neurotechnology-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 0.105$ for algorithm ntechlab_000, giving $FMR(T) = 0.0001$ globally.

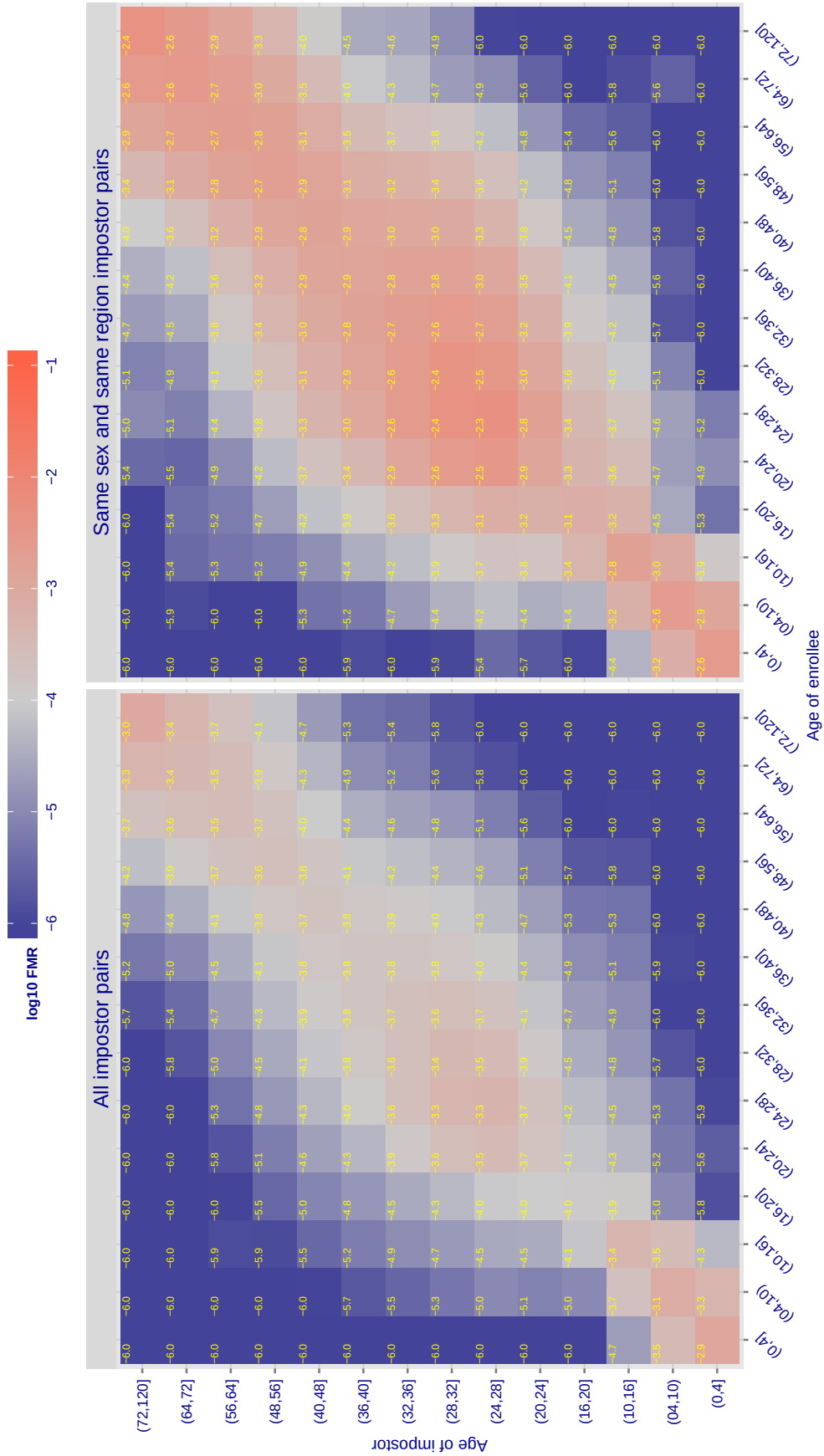


Figure 58: For algorithm ntechlab-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 0.103$ for algorithm ntechlab_001, giving $FMR(T) = 0.0001$ globally.

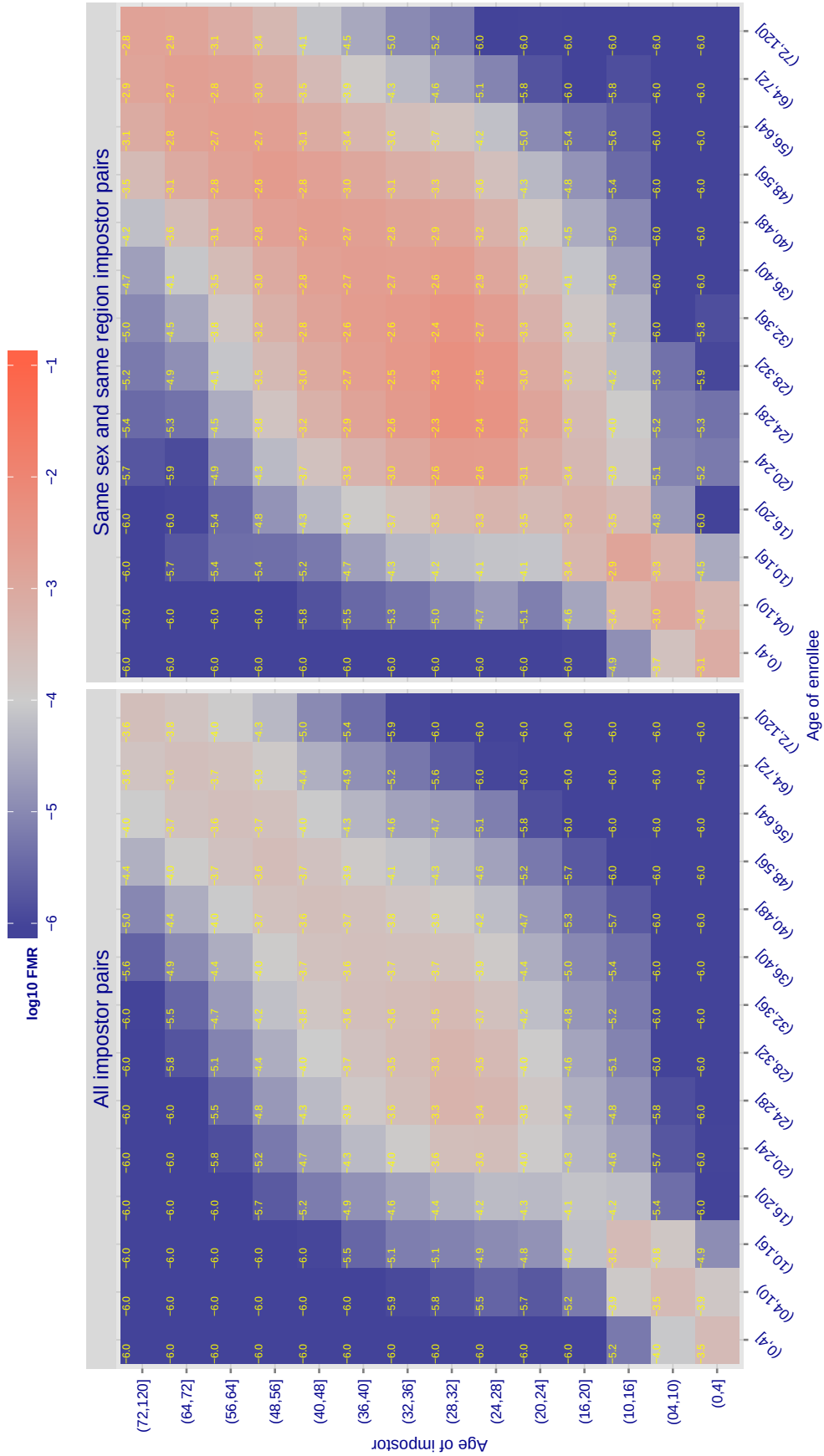


Figure 59: For algorithm ntechlab-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 0.614$ for algorithm rankone_000, giving $FMR(T) = 0.0001$ globally.

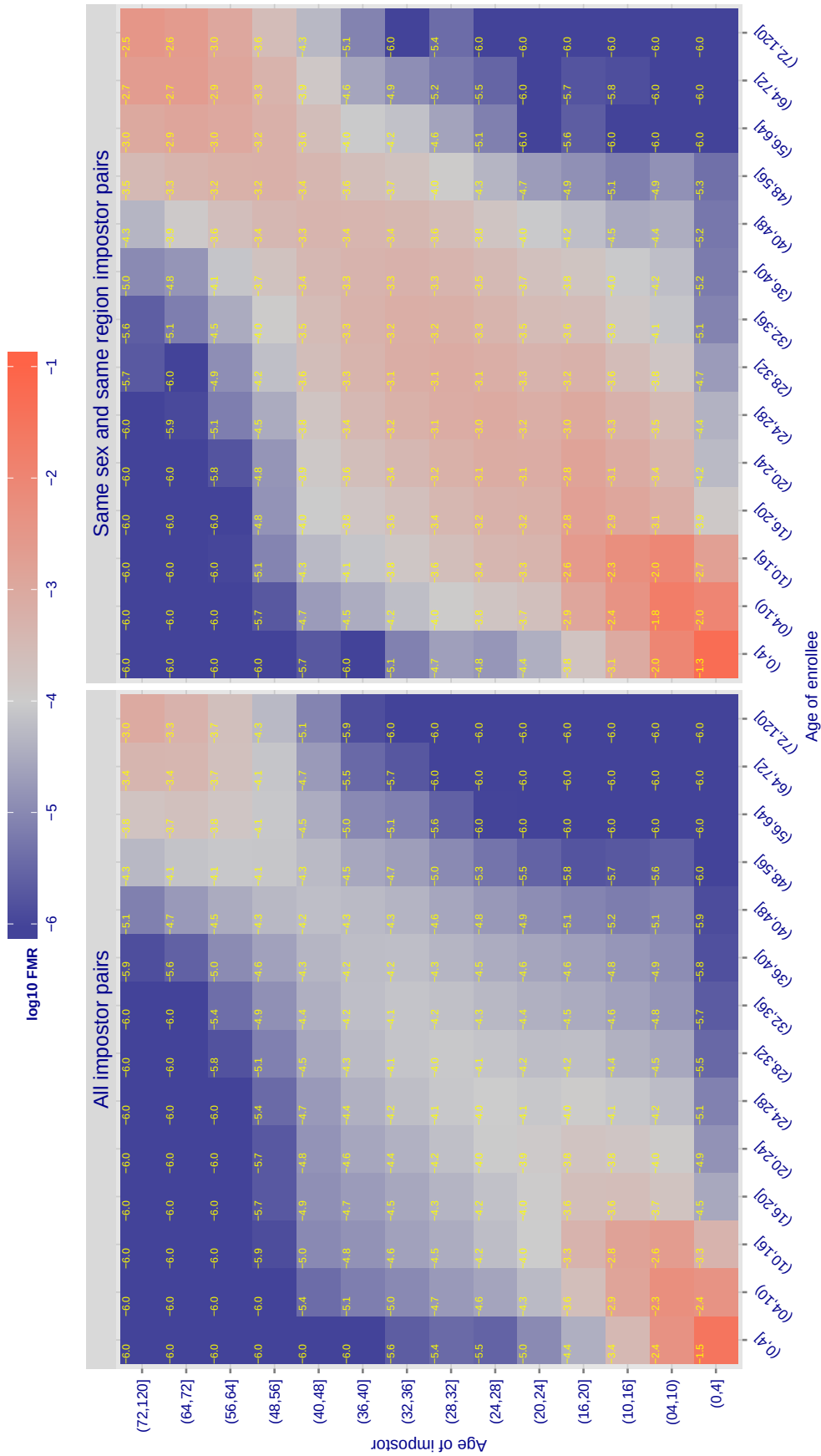


Figure 60: For algorithm rankone-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 0.692$ for algorithm rankone_001, giving $FMR(T) = 0.0001$ globally.

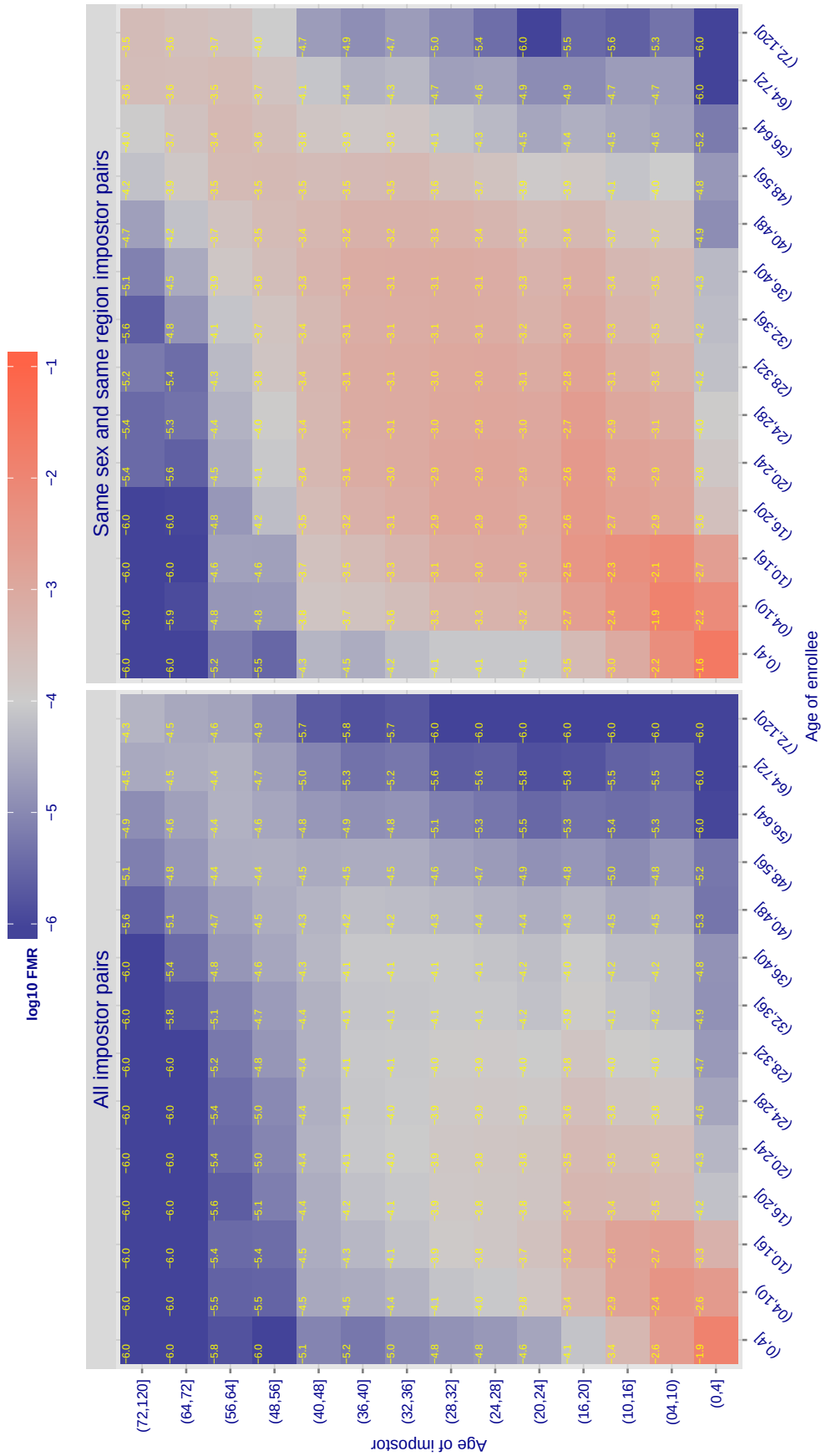


Figure 61: For algorithm rankone-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 80.766$ for algorithm samtech_000, giving $FMR(T) = 0.0001$ globally.

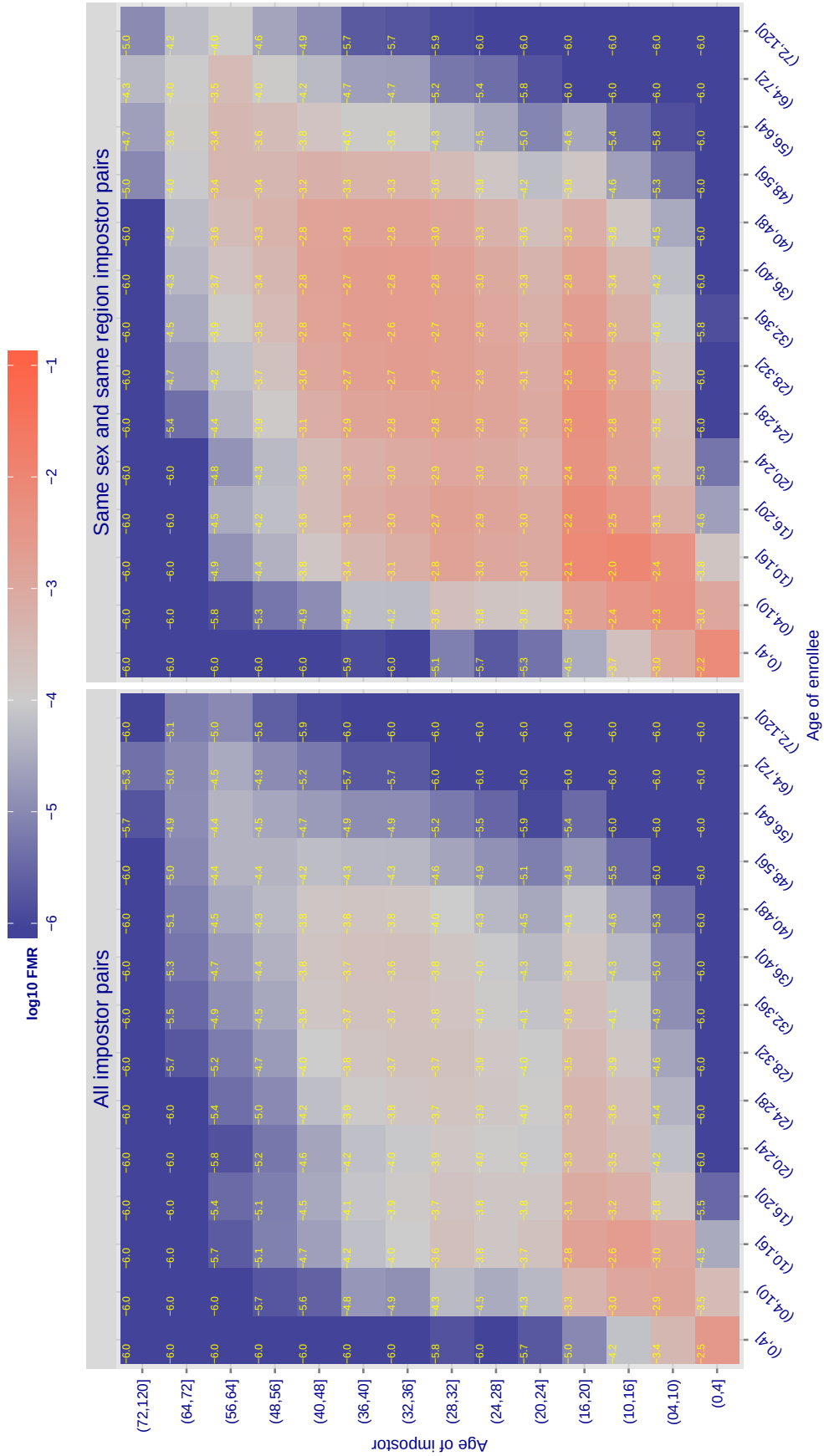


Figure 62: For algorithm samtech-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 10.120$ for algorithm tongyitrans_001, giving $FMR(T) = 0.0001$ globally.

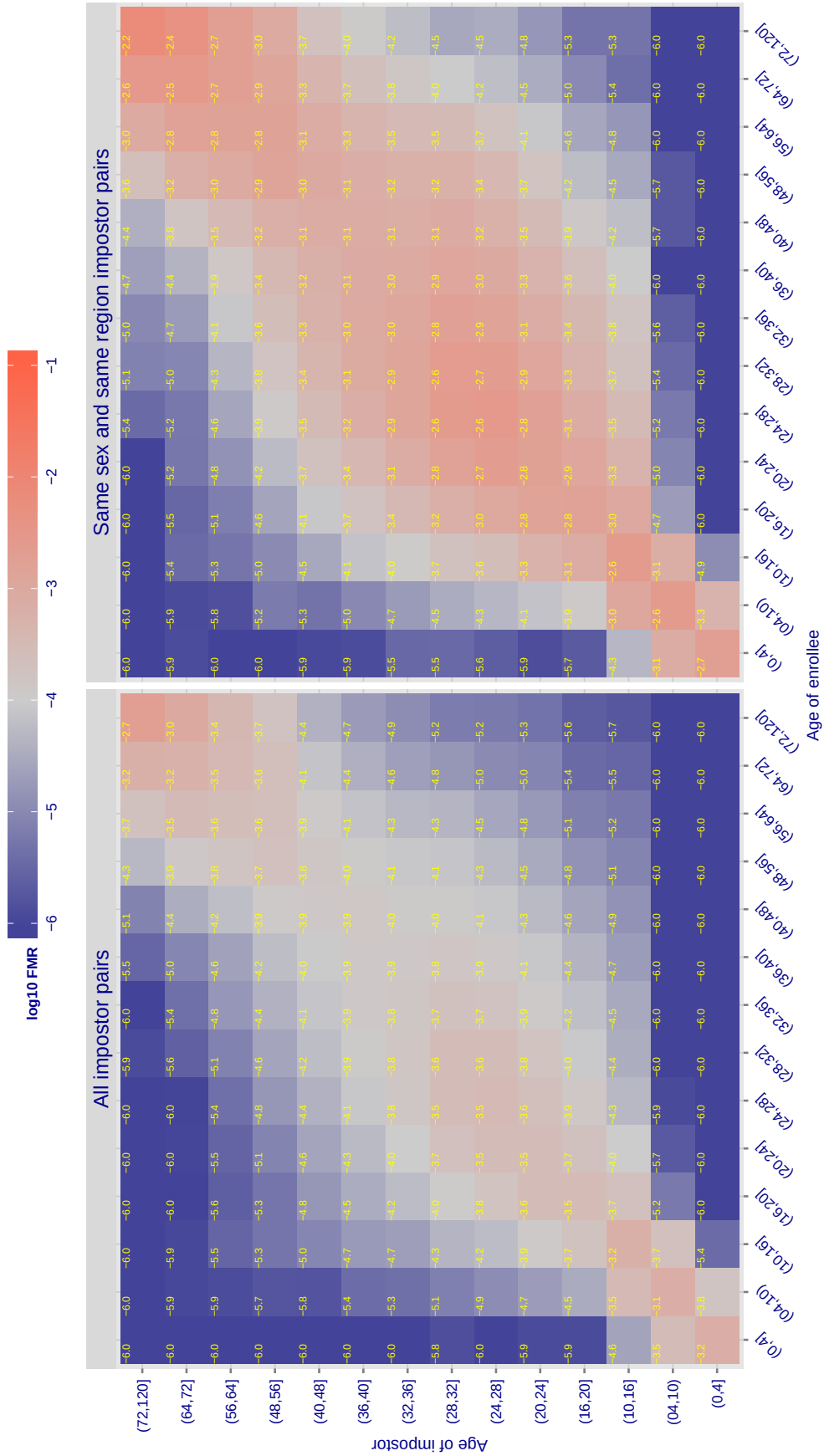


Figure 63: For algorithm tongyitrans-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 1.000$ for algorithm tuple_001, giving $FMR(T) = 0.0001$ globally.

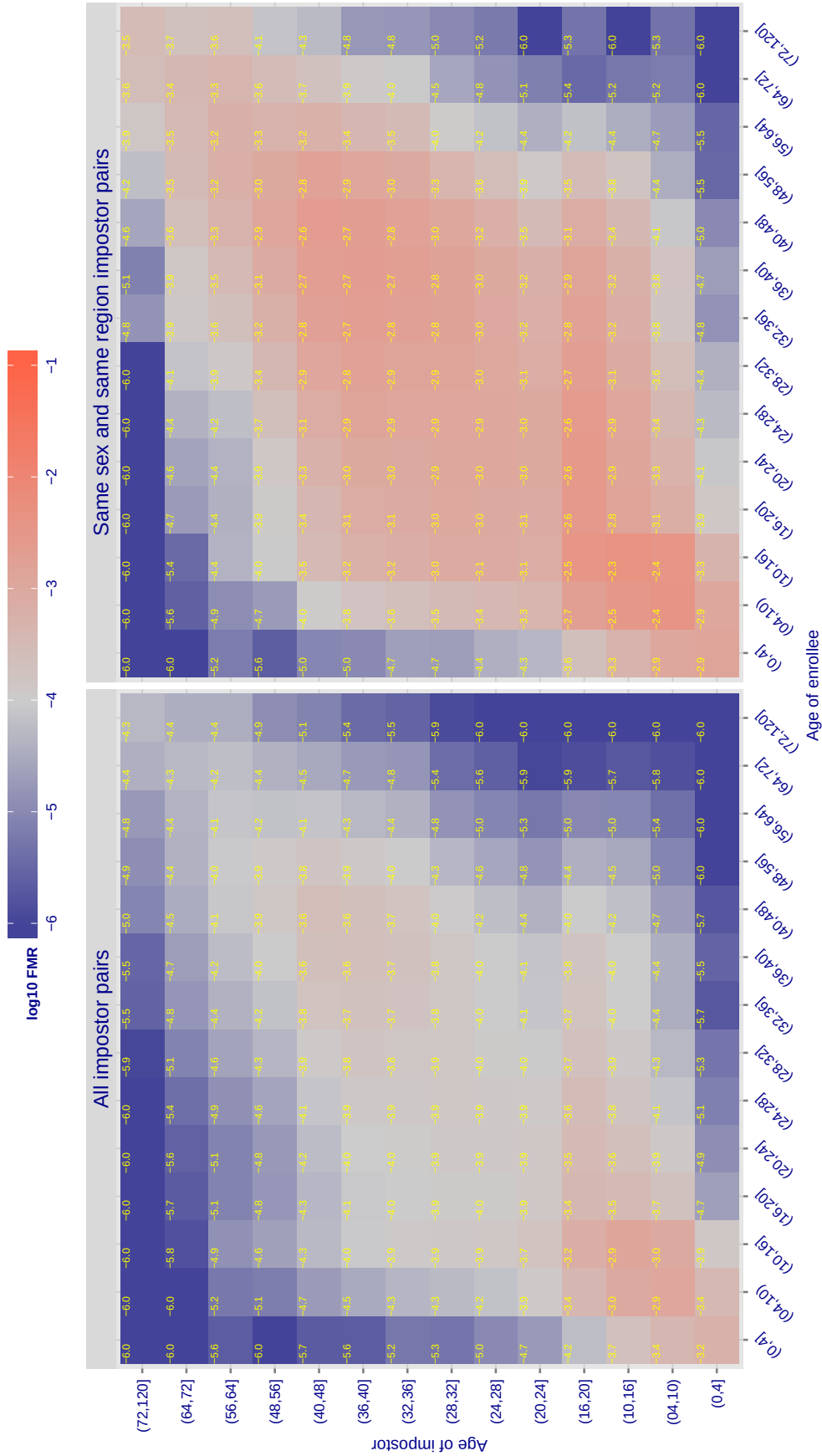


Figure 64: For algorithm tuple_001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 18.505$ for algorithm vcog_001, giving $FMR(T) = 0.0001$ globally.

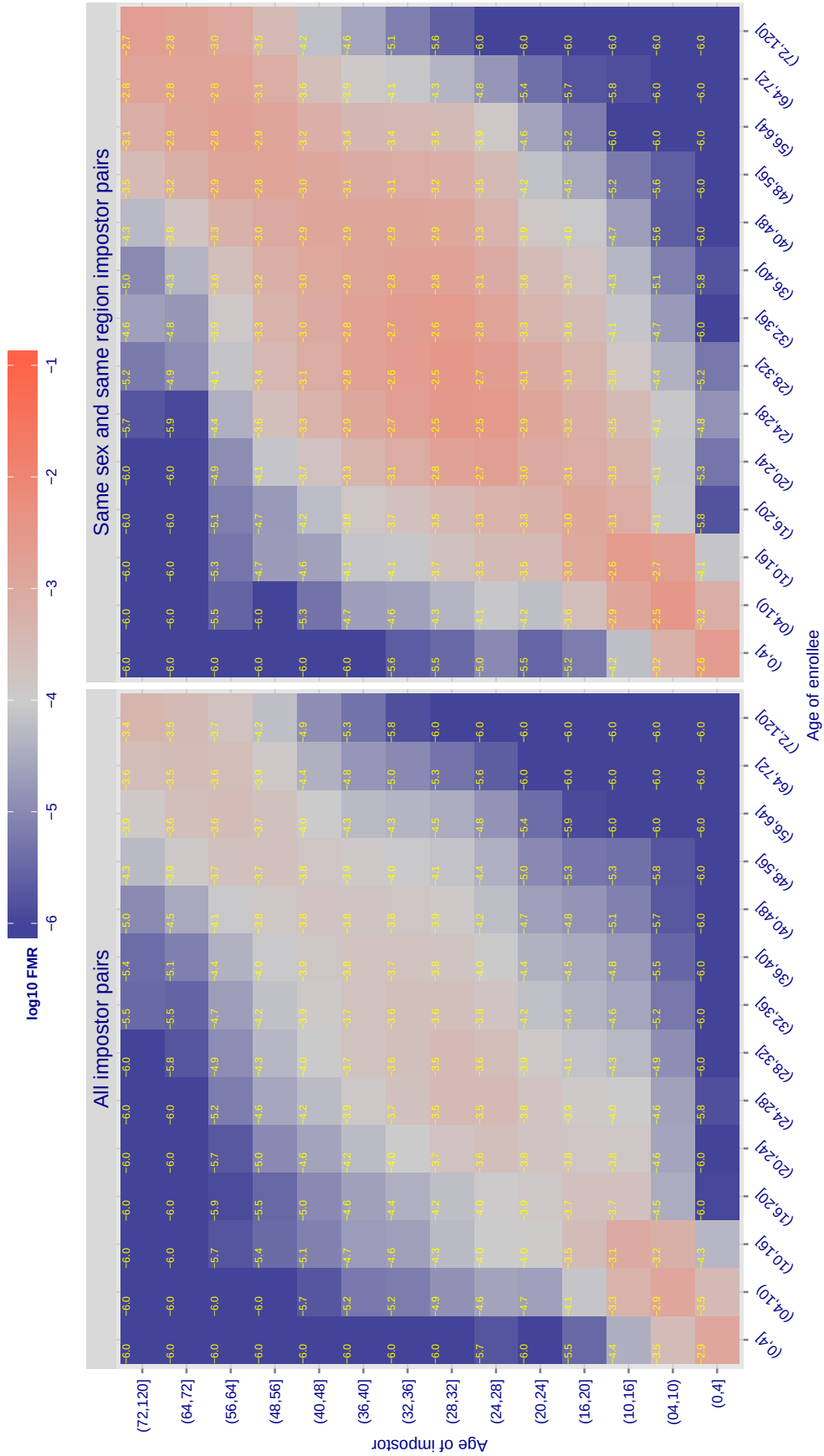


Figure 65: For algorithm vcog-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 0.114$ for algorithm vigilant solutions_000, giving $FMR(T) = 0.0001$ globally.

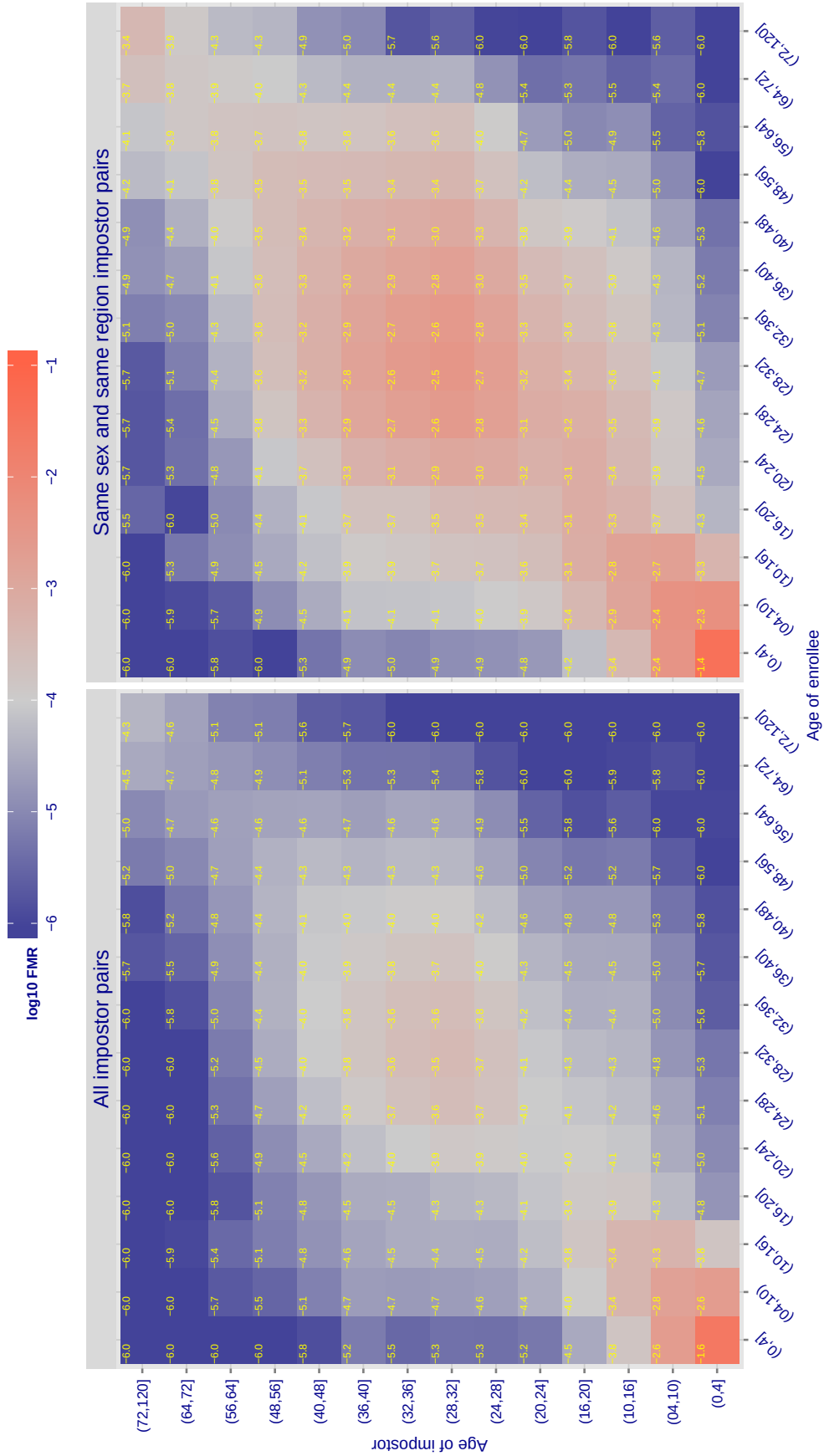


Figure 66: For algorithm vigilant solutions-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 0.080$ for algorithm visionlabs_001, giving $FMR(T) = 0.0001$ globally.

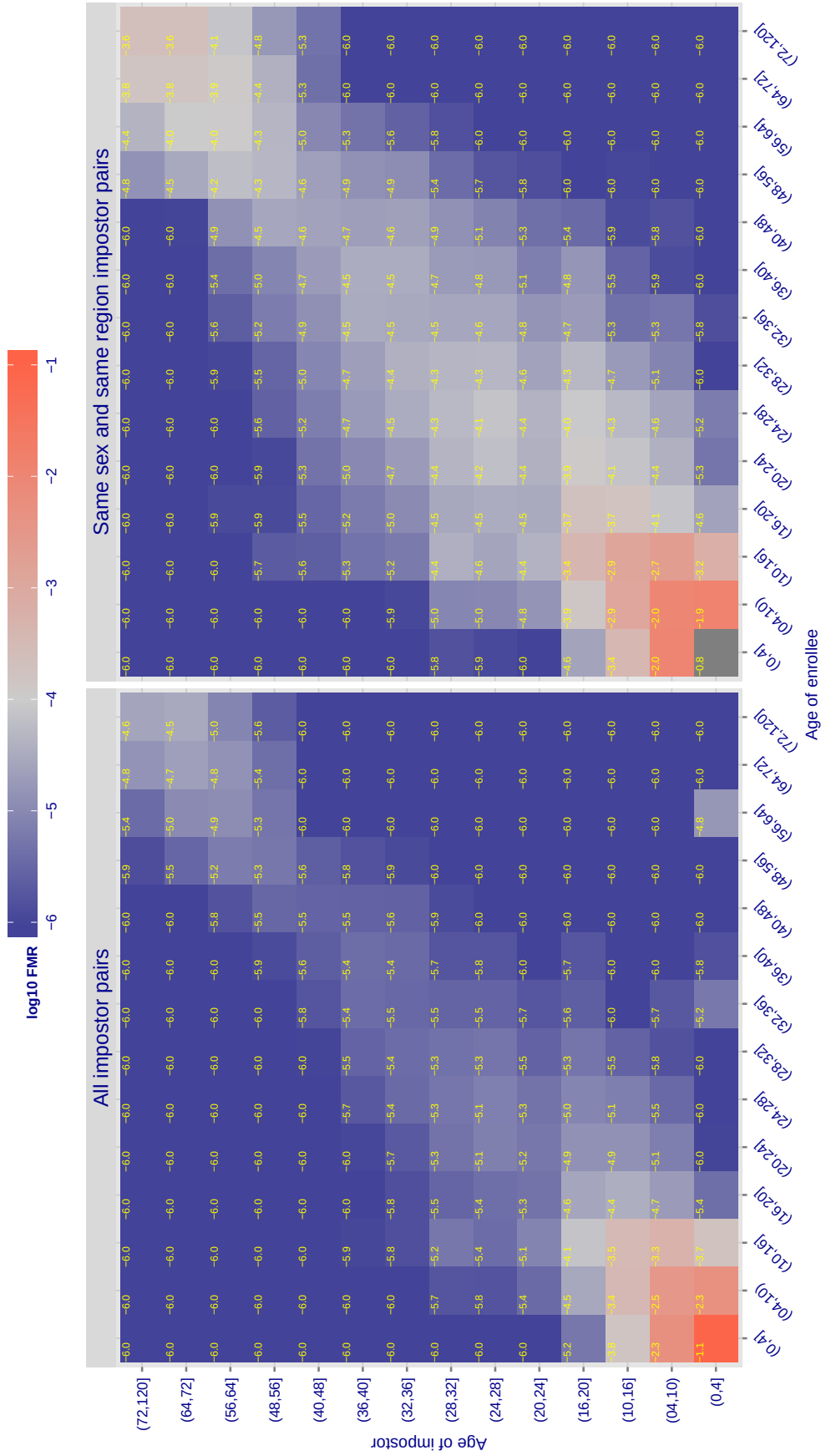


Figure 67: For algorithm visionlabs-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 0.903$ for algorithm vocord_001, giving $FMR(T) = 0.0001$ globally.

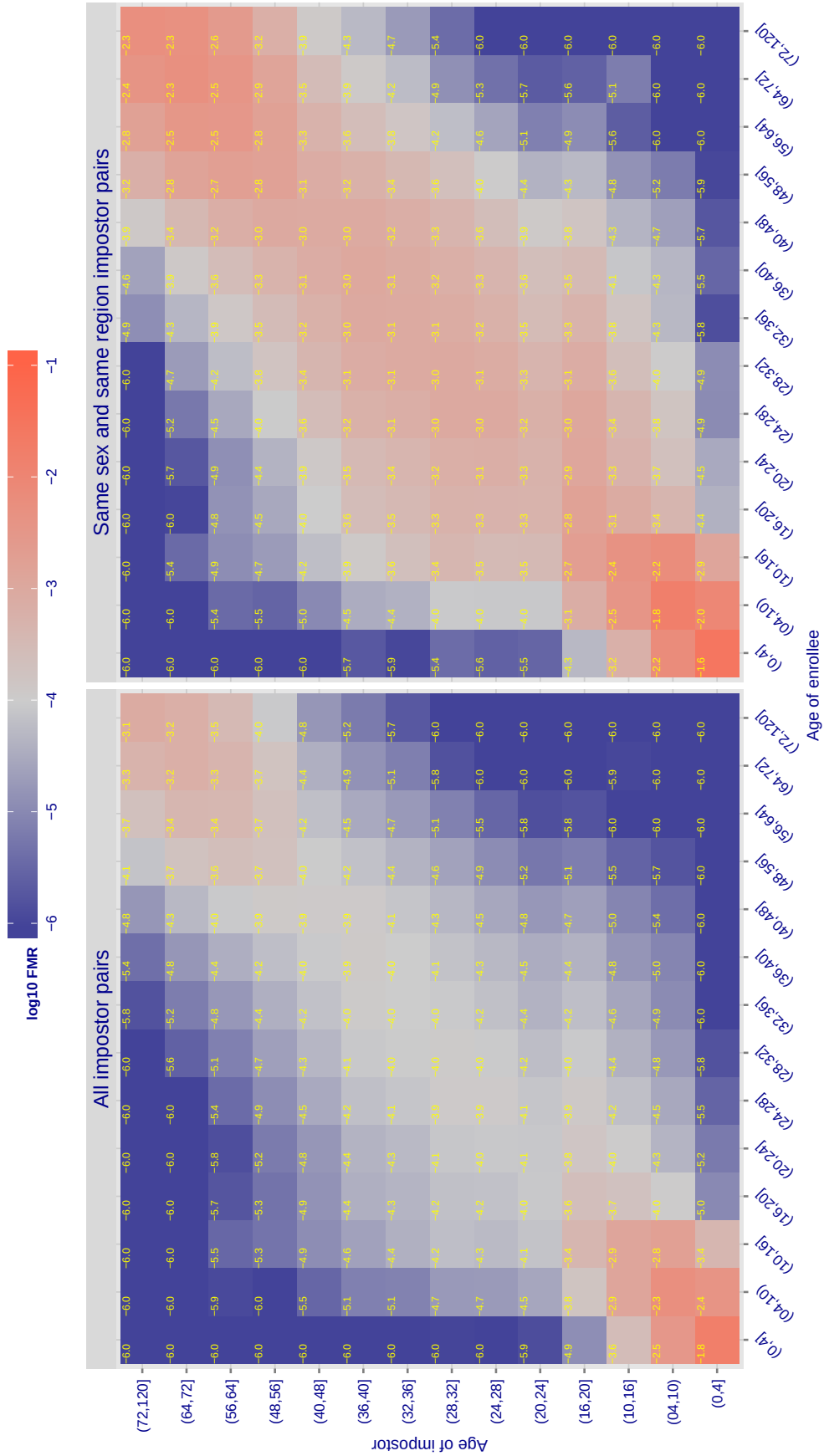


Figure 68: For algorithm vocord-001 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 0.867$ for algorithm vocord_002, giving $FMR(T) = 0.0001$ globally.

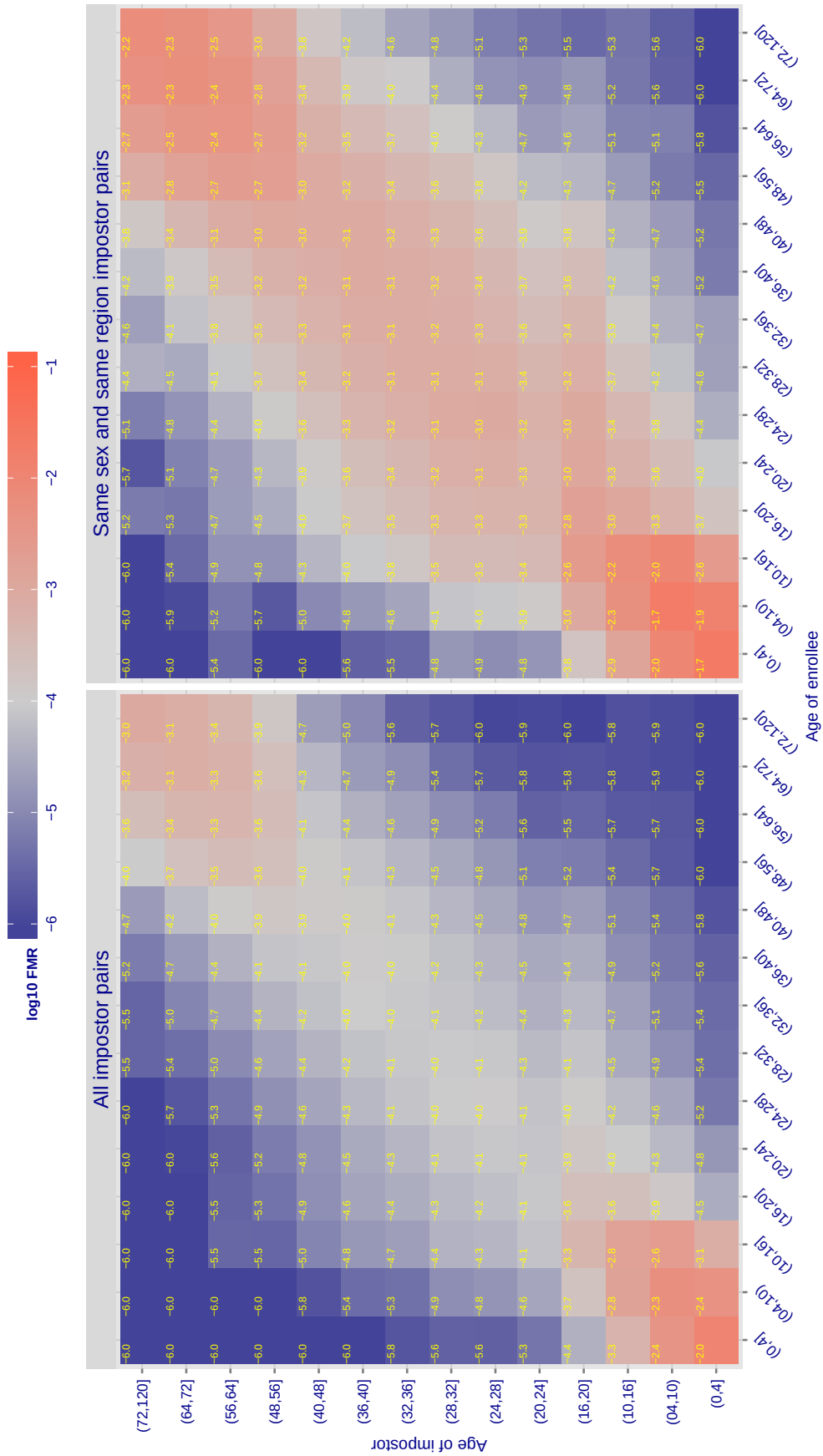


Figure 69: For algorithm vocord-002 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Cross age FMR at threshold $T = 10.098$ for algorithm yitu_000, giving $FMR(T) = 0.0001$ globally.

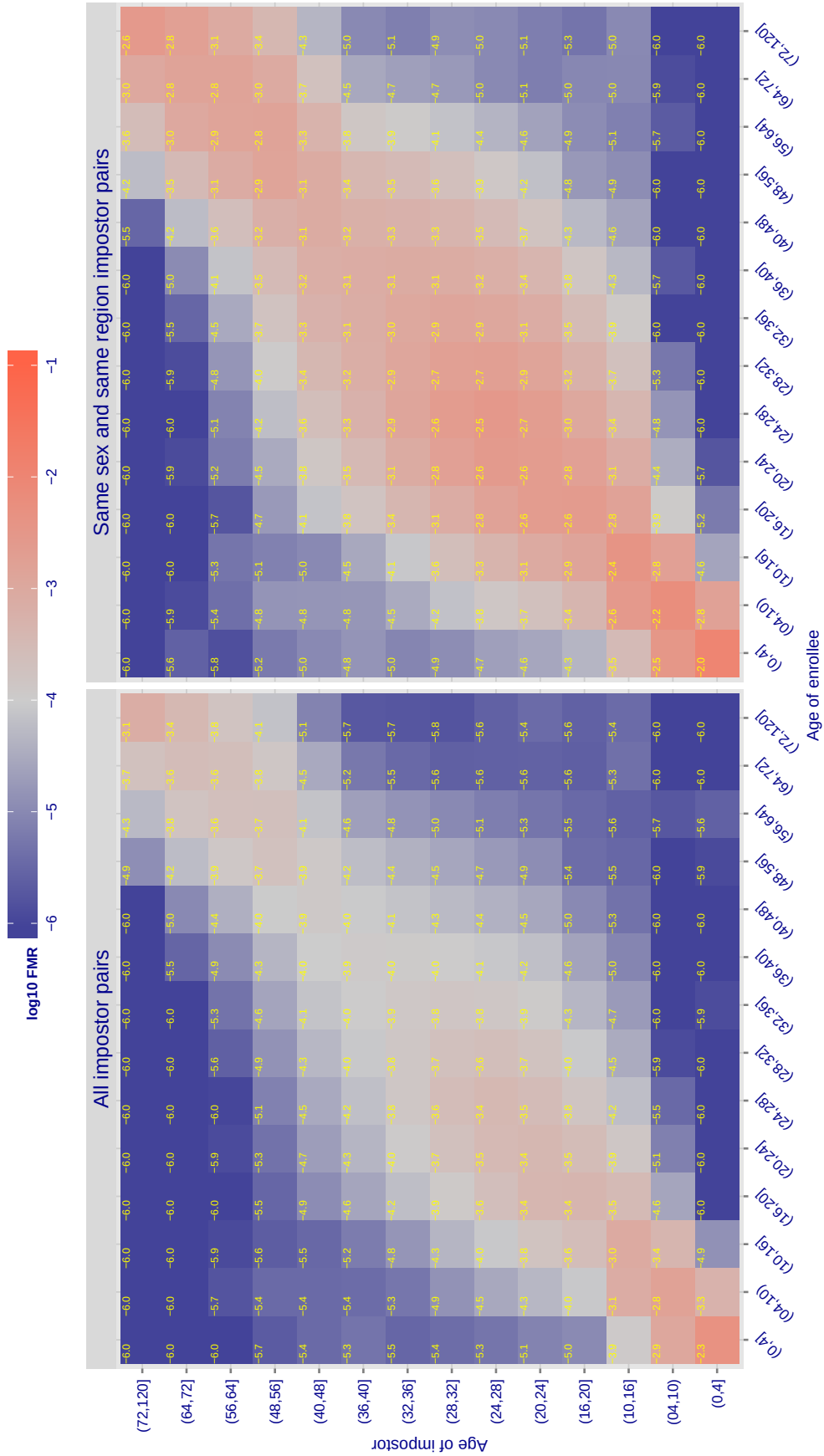


Figure 70: For algorithm yitu-000 operating on visa images, the heatmap shows false match observed over impostor comparisons of faces from different individuals who have the given age pair. False matches are counted against a recognition threshold fixed globally to give $FMR = 0.001$ over all $O(10^{10})$ impostor comparisons. The text in each box gives the same quantity as that coded by the color. Light colors present a security vulnerability to, for example, a passport gate.

Accuracy Terms + Definitions

In biometrics, Type II errors occur when two samples of one person do not match – this is called a **false negative**. Correspondingly, Type I errors occur when samples from two persons do match – this is called a **false positive**. Matches are declared by a biometric system when the native comparison score from the recognition algorithm meets some **threshold**. Comparison scores can be either **similarity scores**, in which case higher values indicate that the samples are more likely to come from the same person, or **dissimilarity scores**, in which case higher values indicate different people. Similarity scores are traditionally computed by **fingerprint** and **face** recognition algorithms, while dissimilarities are used in **iris recognition**. In some cases, the dissimilarity score is a distance; this applies only when **metric** properties are obeyed. In any case, scores can be either **mate** scores, coming from a comparison of one person's samples, or **nonmate** scores, coming from comparison of different persons' samples. The words **genuine** or **authentic** are synonyms for mate, and the word **impostor** is used a synonym for nonmate. The words mate and nonmate are traditionally used in identification applications (such as law enforcement search, or background checks) while genuine and impostor are used in verification applications (such as access control).

A **error tradeoff** characteristic represents the tradeoff between Type II and Type I classification errors. For verification this plots false non-match rate (FNMR) vs. false match rate (FMR) parametrically with T.

The error tradeoff plots are often called **detection error tradeoff (DET)** characteristics or **receiver operating characteristic (ROC)**. These serve the same function but differ, for example, in plotting the complement of an error rate (e.g. $TMR = 1 - FNMR$) and in transforming the axes most commonly using logarithms, to show multiple decades of FMR. More rarely, the function might be the inverse Gaussian function.

More detail and generality is provided in formal biometrics testing standards, see the various parts of [ISO/IEC 19795 Biometrics Testing and Reporting](#). More terms, including and beyond those to do with accuracy, see [ISO/IEC 2382-37 Information technology -- Vocabulary -- Part 37: Harmonized biometric vocabulary](#)

