Modeling lesion transition dynamics to clinically characterize mpox patients in the Democratic **Republic of the Congo**

AUTHORS:

Takara Nishiyama^{1,†}, Fuminari Miura^{2,3,†}, Yong Dam Jeong¹, Naotoshi Nakamura¹, Hyeongki Park¹, Masahiro Ishikane⁴, Shotaro Yamamoto¹, Noriko Iwamoto⁴, Michiyo Suzuki⁴, Ayana Sakurai⁴, Kazuvuki Aihara⁵, Koichi Watashi⁶, William S Hart⁷, Robin N Thompson⁷, Yasuhiro Yasutomi^{8,9}, Norio Ohmagari⁴, Placide Mbala Kingebeni¹⁰, John W. Huggins¹¹, Shingo Iwami^{1,5,12,13,14,15,16,‡,*}, Phillip R. Pittman^{9,‡,*}

AFFILIATIONS:

¹interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan. ²Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands. ³Center for Marine Environmental Studies (CMES), Ehime University, Ehime, Japan. ⁴Disease Control and Prevention Centre, National Centre for Global Health and Medicine, Tokyo, Japan.⁵International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo, Japan. ⁶Research Center for Drug and Vaccine Development, National Institute of Infectious Diseases, Tokyo, Japan. 7 Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK. 8Tsukuba Primate Research Center, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Tsukuba, Japan. ⁹School of Integrative and Global Majors, University of Tsukuba, Tsukuba, Japan.¹⁰Institut National de Recherche Biomédicale, Ministère de la Santé Publique, Democratic Republic of the Congo, ¹¹Division of Medicine, U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID), Fort Detrick, Maryland, United States of America. ¹²Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan. ¹³Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Kyoto, Japan.¹⁴Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Saitama, Japan.¹⁵NEXT-Ganken Program, Japanese Foundation for Cancer Research (JFCR), Tokyo, Japan. ¹⁶Science Groove Inc., Fukuoka, Japan.

†,‡ These authors contributed equally to this study.

^{*}To whom correspondence may be addressed. NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Email: <u>iwami.iblab@bio.nagoya-u.ac.jp</u> (S.I.) and <u>phillip.r.pittman.civ@health.mil</u> (P.R.P)

1 Abstract

2 Coinciding with the global outbreak of clade IIb mpox virus (MPXV), the Democratic Republic of the Congo (DRC) recently experienced a rapid surge in mpox cases with clade I 3 4 MPXV. Clade I MPXV is known to be more fatal, but its clinical characteristics and prognosis differ 5 between patients. Here, we used mathematical modelling to quantify disease progression in a large cohort of mpox patients in the DRC from 2007-2011, particularly focusing on lesion transition 6 7 dynamics. We further analyzed individuals' clinical data to find predictive biomarkers of severity 8 of symptoms. Our analysis shows that mpox patients can be stratified into three groups according 9 to symptom severity, and that viral load at symptom onset may serve as a predictor to distinguish 10 groups with the most severe or mild symptoms after progression. Understanding the severity and 11 duration of symptoms in different patients, as characterized by our approach, allows treatment 12 strategies to be improved and individual-specific control measures (e.g isolation strategies based on disease progression) to be developed. 13

14 Main Text

15 Mpox (formerly known as monkeypox) is a zoonotic disease caused by the monkeypox virus (MPXV), resulting in a rash [1-3]. MPXV was first discovered in humans in the Democratic 16 17 Republic of the Congo (DRC) in 1970 and is classified into two clades (clade I and clade II). Both 18 clades have been endemic in countries in Central and West Africa, spreading mainly by animal-19 to-human transmission [4-8]. A recent global mpox outbreak with clade IIb emerged through 20 human-to-human transmission in Western countries, predominantly affecting men who have sex 21 with men via sexually associated transmission, starting in May 2022 [9-12]. Initially rapid, the mpox outbreak was no longer considered a public health emergency of international concern 22 23 (PHEIC) by May 2023 [13]. Nevertheless, following earlier concerns [14, 15], the number of mpox 24 cases has significantly increased in the Western Pacific Region since April 2023, particularly in 25 China. Moreover, the DRC has recently documented transmission via sexual contact of clade I MPXV [16], which has a case fatality ratio of about 10% [6, 7, 17, 18], significantly higher than 26 27 that of clade IIb during 2022 (<1%) [19-22]. There has been a rapid increase in the number of 28 suspected cases of clade I in the DRC (with more than 10,000 reported cases in 2023) [23], 29 further raising concerns about the potential global spread of mpox, potentially with this more fatal clade, despite termination of the PHEIC due to clade IIb. 30

Understanding disease progression is key for both treatment and outbreak control. 31 Common symptoms of mpox include a skin rash or mucosal lesions accompanied by fever, 32 33 headache, muscle aches, back pain, low energy, and swollen lymph nodes [10, 22, 24-27]. The 34 skin lesions associated with mpox typically persist for 2 to 3 weeks and progress through seven 35 stages: macules, papules, vesicles, and then pustules, followed by umbilication, scabbing, and 36 desquamation [28-30]. Current recommendations for patients are to refrain from direct skin contact from the symptom onset until lesions have healed and scabs fall off (i.e., a symptom-37 based isolation rule [31]), given that person-to-person transmission of mpox can occur through 38 39 direct contact with infectious skin or other lesions [28-30]. Notably, the viral load in skin samples

40 is significantly higher, by several orders of magnitude, and persists longer than the viral load in 41 other body locations [32, 33]. The number of skin lesions that a patient experiences during MPXV 42 infection varies over time and between individuals, ranging from a few to over 1000 [8, 25, 27]. Clinical features of the rash also differ depending on MPXV clade [26, 32, 34, 35]. Thus, 43 44 individual-level dynamics in the number of lesions provide evidence for determining tailored 45 treatment and effective interventions. However, mpox lesion dynamics have not previously been 46 characterized quantitatively, and in particular, individual heterogeneity has yet to be fully 47 examined.

48 In this study, we aimed to quantify the time course of disease progression of mpox and find predictors of individual-level severity of symptoms, using a large longitudinal dataset that 49 contains lesion counts, viral loads of various specimens, and other biomarkers from patients 50 51 infected with clade I MPXV in the DRC during 2007–2011 [27]. Our analysis stratified the study population into three groups by severity of lesion symptoms (e.g., mild or severe progressors), 52 53 clinically characterized them, and identified clinical markers predicting severity. These findings 54 have the potential to contribute to the development of effective treatment strategies and control 55 measures, not only for clade I MPXV but also other MPXV clades and human orthopoxviruses 56 [36].

57 **Results**

58 Description of cohort and study design

We used comprehensive clinical data collected from MPXV-infected patients at the remote L'Hopital General de Reference de Kole (Kole Hospital), in the rainforest of the Congo River basin of DRC, from March 2007 until August 2011. The data were collected as part of a study conducted jointly by the Institute National de Recherche Biomedical (INRB) and the US Army Medical Research Institute of Infectious Diseases (USAMRIID) [27].

Of 244 patients diagnosed with MPXV infection, 228 tested positive by MPXV-specific 64 PCR testing. In addition, 13 patients were excluded because of incomplete observational data. In 65 our analysis, we studied lesion transition dynamics in 149 participants and virus dynamics in 151 66 67 participants, excluding 66 and 64 individuals, respectively, owing to incomplete longitudinal data on lesion counts and viral load (see Extended Data Fig 1). Among the 228 patients who tested 68 69 positive, 3 were fetal cases. The dataset comprised lesion counts at different stages on various 70 areas of the patients' bodies, as well as the MPXV genomic DNA load (i.e., viral load) in peripheral 71 blood, pharyngeal swabs, and scabs from the same individuals. Additionally, we integrated clinical 72 laboratory data from hematology analysis, urinalysis, and clinical chemistry within the same 73 patient cohort; **Table 1** shows the summary statistics of these measured items for all individuals 74 (total) and by the level of disease severity (G1, G2, and G3), which is characterized in the next 75 section. Furthermore, the comprehensive clinical dataset for these patients included annotations 76 for their sex, age, clinical symptoms, clinical signs, and exposure possibilities (details are outlined 77 in [27]).

78 Table 1. Summary of clinical laboratory data of 149 mpox patients for lesion transition analysis

	Total (N=149)	G1 (N=42)	G2 (N=31)	G3 (N=76)	Adjusted P value [†]
Chemistry					
Glucose	93.3 [20.9]	87.3 [16.6]*	95.9 [21.8]	94.6 [22.6]	1.000
Blood urea nitrogen	8.23 5.22	6.18 [2.78]	8.63 4.97	10.1 7.18	0.610
Creatinine	0.72 0.50	0.84 0.80	0.69 0.30	0.62 0.31	1.000
Uric acid	4.60 [1.80]	4.44 [1.40]	4.65 [1.95]	4.71 [1.96]	0.978
Calcium	8.82 [0.80]	8.60 [1.29]	8.87 [0.47]	9.01 [0.52]	1.000
Albumin	2.79 [0.55]	2.89 [0.88]	2.73 [0.38]	2.83 [0.33]	1.000
Total protein	7.72 [0.98]	7.67 [1.33]	7.73 [0.74]	7.76 [0.99]	1.000
Alanine aminotransferase	23.6 [10.7]	21.4 [9.52]	25.2 [11.9]	22.7 [8.33]	1.000
Aspartate aminotransferase	44.5 [22.4]	39.5 [24.8]	44.5 [20.1]	51.3 [23.5]	0.377
Alkaline phosphatase	118 [45.6]	116 [52.1]	117 [40.2]	123 [49.9]	0.942
Total bilirubin	0.62 [0.19]	0.61 [0.20]	0.60 [0.19]	0.66 [0.20]	1.000
Gamma glutamyl transferase	40.6 [36.9]	32.4 [19.3]	42.1 [40.0]	48.0 [45.0]	1.000
Amylase	79.3 [102]	75.1 [40.1]	82.0 [123]	78.3 [100]	1.000
Complete blood count					
Neutrophils	38.8 [13.1]	33.5 [12.8]	39.1 [11.0]	45.2 [15.7]	0.080
BANDS	1.17 [2.98]	0.65 [1.90]	1.12 [2.46]	1.97 [4.72]	1.000
Lymphocytes	41.2 [15.7]	46.7 [13.3]	40.9 [15.9]	34.6 [15.8]	0.210
Atypical lymphocytes	0.78 [1.33]	0.63 [1.25]	0.91 [1.35]	0.68 [1.40]	1.000
Metamyelocytes	0.01 0.08	0 [0]	0.01 [0.12]	0[0]	1.000
Myelocytes	0.27 [0.73]	0.10 [0.38]	0.33 [0.83]	0.36 [0.80]	1.000
Monocytes	5.55 [6.61]	5.30 [6.84]	6.17 [6.91]	4.36 [5.51]	1.000
Eosinophils	11.5 [6.48]	12.8 [6.93]	11.1 [5.89]	10.6 [7.21]	1.000
Basophils	0.29 [0.59]	0.30 [0.46]	0.23 [0.48]	0.42 [0.89]	1.000
White blood cell	10.7 [5.15]	9.29 [4.26]	10.4 [5.13]	13.3 [5.44]	0.017
Red blood cell	4.50 [0.71]	4.45 [0.78]	4.49 [0.67]	4.61 [0.75]	1.000
Hemoglobin	11.6 [2.02]	11.4 [2.22]	11.6 [1.83]	11.8 [2.31]	1.000
Hematocrit	34.7 [5.84]	34.5 [6.33]	34.6 [5.55]	35.1 [6.02]	1.000
Mean corpuscular volume	77.9 [7.72]	79.1 [8.99]	77.6 [7.23]	77.1 [7.25]	1.000
Mean corpuscular hemoglobin	25.9 [2.85]	25.8 [3.24]	25.9 [2.68]	25.6 [2.88]	1.000
Mean corpuscular hemoglobin concentration	33.2 [1.89]	32.6 [1.64]	33.5 [2.08]	33.2 [1.49]	1.000
Platelet	275 [154]	244 [118]	273 [146]	318 [204]	1.000
Malaria smear					
Peripheral blood parasite detection	1 (1)	1 (0.7)**	0 (0)	0 (0)	1.000
Urinalysis					
Urine glucose	17 (11)	5 (3.4)	7 (4.7)	5 (3.4)	1.000
Bilirubin	112 (75)	29 (19.5)	63 (42.3)	20 (13.4)	1.000

Ketones	75 (50)	20 (13.4)	39 (26.2)	16 (10.7)	1.000	
Specific gravity***	110 (74)	34 (22.8)	52 (34.9)	24 (16.1)	1.000	
	26 (17)	5 (3.4)	19 (12.8)	2 (1.3)		
Occult hematuria	31 (21)	8 (5.4)	17 (11.4)	6 (4)	0.963	
Urine pH	6.41 [1.05]	6.59 [1.10]	6.39 [0.97]	6.22 [1.16]	1.000	
Proteinuria	114 (77)	31 (20.8)	62 (41.6)	21 (14.1)	1.000	
Urobilinogen	56 (38)	12 (8.1)	31 (20.8)	13 (8.7)	1.000	
Nitrites	30 (20)	6 (4)	20 (13.4)	4 (2.7)	1.000	
Leukocytes esterase	37 (25)	11 (7.4)	20 (13.4)	6 (4)	1.000	

[†] Statistics for G1 (mild progressors), G2 (intermediate progressors), and G3 (severe progressors) performed by Kruskal-Wallis test and Fisher's exact test with FDR.

* mean [S.D.].
** No. of positive (%).
*** The top row shows the number of people in specific gravity level 2 and the bottom row shows the number of people in specific gravity level 3.

83 Quantifying and stratifying lesion transition dynamics of MPXV-infected patients

Generally, mpox lesions can be classified into seven distinct stages: macules, papules, vesicles, pustules, umbilication, scabbing, and desquamation [27, 28, 32]. As depicted in the distribution of lesion counts at various stages in **Extended Data Fig 2**, the average lesion count at the macule, papule, vesicle, umbilicated, and pustule stages reached zero by approximately 20 days after onset of the first lesion, whereas the count at the scab and desquamation stages persisted even beyond 40 days post-onset (**Fig 1A**). Nevertheless, there was significant individual-level variation in the number of lesions and in the temporal dynamics of lesion transition

91 (Extended Data Fig 2).

To quantify the variation among MPXV-infected patients (e.g., **Extended Data Fig 2**), we 92 93 used a mathematical model that describes the dynamics of lesion transition (see Methods in 94 detail). For model fitting, we merged the macule, papule, and vesicle stages into one stage, owing to the limited number of lesions observed at each stage and the lack of clinical relevance in 95 96 distinguishing between these three stages. Additionally, considering that the infectious period 97 persists until all scabs (or crusts) have fallen off [28], we excluded the desquamation stage from 98 our analysis. Consequently, we modelled lesion counts using the compartments L_1 , L_2 , L_3 , and L_4 , in which L_1 represents lesions in the macule, papule, and vesicle stages, and L_2 , L_3 , and 99 L_{4} represent lesions in the pustule, umbilicated, and scab stages, respectively (**Extended Data** 100 101 Fig 3). We then characterized the transition dynamics among these stages at the individual level. We reconstructed the lesion transition dynamics for each participant in our cohort over a period 102 103 of approximately 40 days following the onset of lesions (see Methods for details). The individualspecific model fits are shown in Fig S1 and the estimated parameters are summarized in Table 104 **S1**. 105

We performed an unsupervised clustering analysis using a dissimilarity-based random forest clustering approach [37, 38] to stratify the time-course patterns of lesion transition based on estimated individual-level parameters (see **Methods** in detail). The number of clusters optimized by the algorithm was three (i.e., three groups of patients; G1, G2, and G3). **Fig 1B**

110 represents a two-dimensional Uniform Manifold Approximation and Projection (UMAP) 111 embedding of these three groups, which clearly shows that G1 is distinct from the other groups. Using a different color for each group, we also plotted the reconstructed individual transition 112 dynamics in **Extended Data Fig 4**. The corresponding projected time course of each group in **Fig** 113 **1C** shows that the lesion transitions between stages progressed rapidly among individuals in G1 114 115 (N=42), whereas they were slower in the G3 participants (N=31). On the other hand, individuals in G2 (N=76) showed intermediate patterns, even though the dynamics of the scab disappearance 116 stage (the cyan curve) were comparable to that in G3. These patterns suggest substantial 117 118 variation in the infectious period of MPXV-infected patients, consistent with our recent findings of 119 substantial individual variations in the duration of viral shedding in different specimens of MPXV clade IIb B.1 [39]. Our statistical tests showed that individuals in G1 and G3 had significantly 120 smaller and larger numbers of lesions for all stages, respectively, compared with individuals in 121 the other groups (Fig 1D). Additionally, we observed statistically significant differences in total 122 123 lesion counts and lesion duration (defined as time between onset and lesion disappearance) between patient groups, except between G2 and G3 in lesion duration (Fig 1E). These findings 124 125 demonstrate that G2 and G3 are distinguished by the total number of lesions as well as by the 126 lesion counts in each stage, rather than by lesion duration. Hereafter we classify G1 as "mild 127 progressors" and G3 as "severe progressors" in our analysis.

To describe the difference in timing of lesion disappearance, we calculated the probability 128 129 of detectable lesions over time by using the model with estimated parameters for each group (Fig **1F**). We found that individuals in G1 had significantly shorter periods of lesion symptoms: there 130 was a difference in lesion disappearance between G1 vs G2, and between G1 vs G3 (p<0.0001 131 by log-rank test for both). In the total group (i.e., a group of all analyzed cases), the probability of 132 detectable lesions was 55% (95% CI: 48%-64%) at 31 days after lesion onset, consistent with a 133 reported median duration of experiencing apparent lesions among mpox cases in the 2022 134 outbreak [40]. All lesions in G1 patients had disappeared at 31 days after lesion onset, whereas 135 the corresponding probabilities in G2 and G3 were 51% (95% CI: 41%-64%) and 71% (95% CI: 136

137 57%-89%), respectively. In all stratified groups, the probability of detectable lesions was less than
138 10% by 44 days after lesion onset.

139 Additionally, we compared our stratified groups with two major categorizations of MPXVinfected patients to assess reliability from a clinical standpoint: "CDC [Centers for Disease Control 140 141 and Prevention lesion severity categories" [25] based on the total number of lesions on admission 142 day (i.e., individuals with <25, 25-99, 100-499, and >=500 lesions are categorized as 1, 2, 3, and 4, respectively) and "age categories" [27] (individuals <5, 5-11, and >=12 years old are 143 categorized as 1, 2, and 3, respectively) (Fig 1G). Both categories were based on clinical 144 145 investigations. We observed that CDC severity categories generally aligned with our stratified 146 groups (e.g., patients in G1 typically had fewer lesions at admission, and those in G3 had more lesions). In our logistic regression analysis, using the number of lesions at symptom onset as the 147 predictor and the stratified group as the outcome, we obtained high ROC-AUC values: 94% for 148 G1 and 82% for G3. Maximizing the Youden Index from the ROC curves, we calculated cutoff 149 values of 107 and 281 lesion counts with 62% and 61% specificity and 87% and 72% sensitivity 150 151 for predicting mild and severe progressors, respectively. Notably, the cutoff value for G1 (i.e., 107 152 lesion counts) closely matched the boundary between CDC lesion severity categories 2 and 3 153 (100 lesion counts) [28]. In terms of age categories, children under 5 years old generally exhibited 154 a longer lesion duration and a larger number of lesions, which aligns with previous reports indicating severe lesion symptoms in young children [17, 27, 41]. 155





157 Figure 1 | Stratification and characterization of lesion transition dynamics: (A) Time course 158 plot of the average lesion count for each stage (i.e., macule, papule, vesicle, pustule, umbilicated, scab, and desquamation) across all patients. (B) UMAP of stratified lesion transition dynamics 159 based on the estimated parameters. Data points represent individual participants and are colored 160 according to group (i.e., G1, G2, and G3: green, blue, and red). (C) The time-course patterns 161 highlighted by Partial Least-Squares Discriminant Analysis (PLS-DA), which discriminates each 162 group from the others. All lesion count values for each time were reconstructed to highlight the 163 difference between lesion stages. The mean (solid line) and mean ± 1 standard deviation (shaded 164 area) calculated based on these reconstructed values are shown, colored according to stage (i.e., 165 L_1 , L_2 , L_3 and L_4 : orange, magenta, light purple, and cyan). (D) Total number of lesions at each 166 stage for each stratified group, calculated based on the reconstructed lesion transition dynamics. 167 (E) Number of lesions (left) and period from lesion onset to disappearance (right), calculated 168 based on reconstructed lesion dynamics. For (D) and (E), statistical significance was calculated 169 using the pairwise Mann-Whitney U test. Also, p-values were corrected by Bonferroni's method 170 (NS.: p-value > 0.05, *: p-value \leq 0.05, **: p-value \leq 0.01, and ***: p-value \leq 0.001, 171 respectively). The statistical significance of the difference is indicated on the top of the plots. (F) 172 173 Kaplan-Meier survival curves of lesion disappearance for each group based on reconstructed lesion dynamics. Solid lines and shaded areas indicate the survival probabilities and 95% 174 175 confidence intervals, respectively. The green, blue, red, and black colors correspond to G1, G2. G3, and the total group, respectively. The table below the figure shows the number and 176 177 percentage of survivors in each group at each time point. (G) Confusion matrices on the relationship between the stratified group and CDC severity score (top panel, individuals with <25, 178 25-99, 100-499, and >=500 lesions) as well as between stratified group and age category (bottom) 179 panel, individuals <5, 5-11, and >=12 years old). 180

181 Viral load at lesion onset may define severity on lesion symptoms

182 While the CDC lesion severity categories are useful, it is practically (and ethically) challenging in clinical settings to always count all lesions appearing on the entire body of a patient 183 [42]. Recent studies have suggested that the diagnosis of small lesions is difficult even for 184 clinicians and guite subjective [43]. To provide more objective support for the mpox diagnostics. 185 186 we explored the relationship between the lesion transition dynamics and individual-level viral load, as our recent study suggested that viral load in lesions is positively associated with the infectious 187 period of a patient (i.e., the duration of having culturable MPXV) for clade IIb MPXV [33, 34, 44]. 188 189 In our cohort, although we measured the viral load in peripheral blood, pharyngeal swabs, 190 and scabs [27] (see Methods), our focus was on the viral load in peripheral blood because the 191 other viral loads are not routinely measured (typically once per patient in most cases). All patients (N=151; Extended Data Fig 1) showed a decaying viral load profile (Fig 2A), since the viral load 192 is likely to have already peaked before the first measurement (i.e., before patients present to the 193 194 hospital) [27]. Therefore, we used a decay model to reconstruct the best-fit viral load profile for each case, taking into account inter-patient heterogeneity (see Methods). In Fig 2B, we depict 195 the reconstructed individual dynamics aligned with the stratified groups (i.e., G1, G2, and G3). 196 197 We compared four key features derived from the viral dynamics among the groups in Fig 2C: viral 198 load at lesion onset and at disappearance (when scab count falls below 1), total viral load (defined as the AUC of viral load above 10^{-2} genomes/mL), and the viral clearance rate. Interestingly, 199 200 we found that the onset viral loads and total viral load in G1 were significantly lower than those in G3, whereas the viral loads at lesion disappearance in G1 were significantly higher than those in 201 G2. Conversely, there was no apparent difference in the clearance rate among the groups. On 202 203 an individual patient level, we observed positive correlations for the two main quantified lesion dynamics features (lesion duration and total lesion counts in Fig 1E) with the onset viral load in 204 205 **Fig 2D** (0.39 and 0.49 Pearson's correlation coefficient, respectively, with a p-value of 0.001 in both cases). These findings suggest that the onset viral load may serve as a potential biomarker 206 for predicting lesion transition patterns and severity of illness over the course of infection 207

208 (discussed below). Additional correlation analyses between lesion dynamics and viral load
 209 dynamics are illustrated in Extended Data Fig 5.

Finally, we explored whether an individual's risk group (G1, G2, or G3) could be predicted 210 by the onset viral load in peripheral blood. We used a logistic regression analysis with the onset 211 212 viral load as the predictor variable and the stratified groups as the outcome variable. Although we 213 could not attain a high ROC-AUC for predicting G2 (41%), which exhibits intermediate patterns, we did achieve relatively higher ROC-AUCs of 69% for G1 (defined as mild progressors) and 67% 214 for G3 (defined as severe progressors) as shown in Fig 2EF. Similarly, based on the Youden 215 216 Index, we calculated cutoff values for the onset viral load of 4.557 and 4.619 (log₁₀ genomes/mL) with 61.8% and 61.1% specificity and 87.1% and 72.0% sensitivity for predicting mild and severe 217 progressors, respectively. Interestingly, these two cutoff values were very close, and therefore 218 219 we henceforth use a common threshold of an onset viral load of 4.6 (log₁₀ genomes/mL) instead. Fig 2G shows that the threshold generally distinguishes between G1 and non-G1 (i.e., mild or 220 221 non-mild progressors) as well as between G3 and non-G3 (i.e., severe or non-severe progressors), but not the intermediate pattern (i.e., G2). Notably, the threshold viral load 222 significantly differentiated the duration of lesions (p<0.0001 by Mann-Whitney U-test): 28.0 and 223 224 35.6 days on average are required until lesion disappearance for patients with an onset viral load 225 below and above 4.6 (log₁₀ genomes /mL), respectively (Fig 2H).



227 Figure 2 | Relationship between lesion transition dynamics and peripheral blood viral load: 228 (A) The measured individual decay dynamics of mpox viral load (i.e., genomic DNA) in peripheral blood samples obtained from the DRC cohort. (B) The individual-level mpox decay viral load, 229 reconstructed using the mathematical model, plotted aligned with the stratified groups (i.e., G1, 230 231 G2, and G3: green, blue, and red). (C) Comparison of viral load at onset, viral clearance rate, 232 total viral load in peripheral blood, and viral load at lesion disappearance among the stratified 233 groups. (D) Correlations between the features of individual-level lesion transition dynamics (i.e., 234 the lesion duration and total lesion counts) and those of individual-level decay viral load dynamics (i.e., the viral load at lesion onset and disappearance). (E)(F) The ROC curves for a generalized 235 linear model (GLM) logistic regression with the viral load at lesion onset for predicting stratified 236 237 group G1 or not and G3 or not, respectively. The corresponding ROC-AUCs are displayed in the 238 bottom right of each panel. FPR and TPR mean false positive rate and true positive rate, 239 respectively. (G) Swarm plot of the viral load at onset among the stratified groups. The dashed line indicates the threshold value (4.6 log₁₀ genomes/mL). (H) Comparison of lesion duration 240 between two groups classified based on the threshold value (Low: below 4.6 log₁₀ genomes/mL. 241 242 High: above 4.6 log₁₀ genomes/mL).

243 Characterizing mild and severe progressors by clinical laboratory data

244 In this section, we further explored potential biomarkers for use in predicting mild and 245 severe progressors by incorporating additional clinical laboratory data. Using the clinical laboratory data obtained from our cohort [27] (see **Methods**), we annotated the 149 patients with 246 hematology analysis, urinalysis, and clinical chemistry on the day of lesion onset (Table 1 and 247 248 **Extended Data Fig 1**). We examined the correlation between the stratified group based on 249 estimated lesion transition patterns and the 41 variables (summarized in **Table 1**). Each factor was compared between the three groups by Kruskal-Wallis and Fisher's exact test where the p-250 values were corrected by the false discovery rate (FDR). However, we found that only the white 251 252 blood cell count differed significantly among the stratified groups (i.e., adjusted p-value of less 253 than 0.05).

Next, we investigated whether the inclusion of the clinical laboratory data alongside the 254 onset viral load improved the performance of the model predictions in **Fig 2EF**. We here used a 255 256 supervised machine learning algorithm (Light Gradient Boosting Machine [LightGBM]) to predict the group from the laboratory data of 120 patients in our cohort (see details in **Methods**). Despite 257 the inclusion of additional data, we achieved similar ROC-AUCs, specifically for G1 (70%) and 258 G3 (64%), as depicted in (Fig 3AB), along with a lower ROC-AUC for G2 (53%). Interestingly, 259 the combination of viral load data and clinical laboratory information did not significantly improve 260 261 predictability for either mild or severe progressors.

Finally, concentrating on G1 (mild progressors) and G3 (severe progressors), we 262 263 examined features (other than the onset viral load) that predict which of these two nonintermediate groups' individuals belong to. In Fig 3AB, we selected features for our model to 264 265 enhance classification accuracy and efficiency by eliminating redundancy. Specifically, the feature selection was based on the importance of SHapley Additive exPlanations (SHAP) values 266 267 [45], which were added to the LightGBM model in order of significance (see Methods). The features with the highest ROC-AUC in the training dataset were then chosen. Following feature 268 selection, the frequencies of the features selected over 50 iterations [46] for G1 and G3 are shown 269

270 in Fig 3CD. Additionally, the distribution of corresponding features between G1 as opposed to 271 non-G1, and G3 as opposed to non-G3, is depicted in **Extended Data Fig 6**, respectively. The 272 frequency lists of selected features for both G1 and G3 included the onset viral load as the 1st and 8th feature, respectively. This again supports the importance of onset viral load in predicting 273 274 the severity of lesion symptoms. Regarding other biomarkers, we found that low viral load and 275 low AST (which indicates absence of liver damage at onset) were weakly associated with mild progressors. In contrast, for the severe progressors, the most relevant biomarkers were high WBC 276 and AST, which implies elevated inflammatory response and liver dysfunction at onset. These 277 278 clinical findings suggest that inflammatory response and virus-induced organ damage at onset 279 may be linked to disease progression.



280

281 Figure 3 | Clinical characterization of stratified groups: (A) and (B) depict the ROC curves of LightGBM classifiers trained to predict either G1 (i.e., mild progressor) or not, and G3 (i.e., severe 282 progressor) or not, respectively. The corresponding ROC-AUCs are calculated and displayed in 283 the bottom right of each panel. (C) and (D) show bar plots of how often a feature was selected 284 among seeds as an indicator of selection stability. Selected features are shown with a frequency 285 greater than 0.25. OnsetVL: viral load at lesion onset, MCHC: mean corpuscular hemoglobin 286 concentration, GLU: blood glucose level, UA: uric acid, PMN: neutrophils, MCH: mean 287 corpuscular hemoglobin, MCV: mean corpuscular volume, AST: aspartate aminotransferase, 288 LYM: lymphocytes, MONO: monocytes, WBC: white blood cell, EOSIN: eosinophils, GGT: 289 290 gamma glutamyl transferase.

291 **Discussion**

292 In the present study, using a large cohort of mpox cases in the DRC from our previous study [27], we have further investigated the dynamics of lesion symptoms at the individual patient 293 level. We have identified groups of mild and severe progressors (G1 and G3); the G1 group was 294 295 characterized by a lower lesion count and more rapid disappearance of lesions, whereas the G3 296 group exhibited a higher lesion count and slower disappearance of lesions. Our stratification also identified age as a key demographic characteristic affecting lesion dynamics, in accordance with 297 previous studies which suggested that younger age is associated with more likely severe 298 299 outcomes, which determine the necessity and benefits of hospital-based care [17, 27, 41].

Limited research has been done on potential biomarkers for the severity of mpox, which 300 301 could be used to monitor clinical progress and tailor treatment plans in individual patients [47]. In 302 a recent study [48], some peptides were identified to be correlated with severity, although the plasma proteome analysis was confined to a small number of mpox patients. In contrast, our 303 304 analysis with data from over 150 mpox patients suggested that the viral load in peripheral blood at lesion onset may serve as a biomarker capable of predicting severity in the early stages of 305 lesion symptom progression. Specifically, it may be possible to differentiate between mild and 306 severe progressors by comparing their viral load at lesion onset with a threshold value of 4.6 307 308 (log₁₀ copy/mL) and also to predict the possible timing of lesion disappearance. To swiftly prioritize 309 patients for either home treatment or hospitalization, especially for individuals who go on to experience complications [36], the onset viral load has practical advantage, as viral load data can 310 311 be obtained during routine practice in clinical settings, such as when confirming a case with PCR 312 testing. In the future, this research may lead to remote diagnosis of Mpox infections in areas with 313 difficult access to medical care, where prognosis and treatment methods can be determined 314 based solely on images of skin lesions.

Furthermore, adaptive treatments that account for the stratified time-course patterns of lesion transition become feasible if patients can be classified into severity groups shortly after being diagnosed with MPXV infection (i.e., the date of symptom onset). For patients classified in

group G1 (i.e., mild progressors), tecovirimat may represent an appropriate treatment, whereas for those in G3 (i.e., severe progressors), combination treatment using tecovirimat with cidofovir, brincidofovir, or vaccinia immune globulin intravenous (VIGIV) may be reasonable [49, 50]. Establishing an early treatment plan is crucial for improved patient outcomes and efficient health care delivery.

323 Our study has several limitations. First, our analysis relied on patient data from the DRC cohort of cases infected with clade I MPXV, and thus caution should be taken when using our 324 findings to inform treatment or interventions during outbreaks of other clades or orthopoxviruses, 325 326 such as the dominant clade MPXV in 2022 (i.e., clade IIb). Delays between infection and recognizable symptom onset, or between symptom onset and confirmation/hospitalization may 327 differ by clade/virus [51] (and potentially by mode of transmission [52, 53]), and timeliness is 328 crucial when measuring onset viral load data. Overall, our estimates of disease progression after 329 lesion onset were mostly consistent with epidemiological findings observed in 2022 [3, 26, 34], 330 331 and the added value of our analysis was to characterize the individual-level disease progression and quantify the duration of symptoms by the level of severity. Our results also provide timely 332 333 evidence for the control of the clade I mpox outbreaks currently surging in the DRC and 334 surrounding countries [54], which have a higher case fatality ratio than outbreaks due to clade IIb 335 [6, 7, 17, 18]. Second, further investigation is needed on the association between lesion counts 336 at each stage and the mpox viral load in peripheral blood, both in patients who have not been 337 treated with antiviral drugs and in those receiving medications such as tecovirimat, cidofovir, and brincidofovir [49, 55]. This is particularly relevant in outbreak settings where these drugs may be 338 administered to infected individuals and used prophylactically. For example, although tecovirimat 339 is currently available only under emergency authorization, clinical trials assessing its efficacy in 340 the treatment of mpox are underway [56]. In fact, case studies of patients treated with tecovirimat 341 342 show anecdotal improvement in symptoms and viral clearance [57, 58]. Finally, in our machine learning analysis, clinical laboratory data were available for only 120 patients because of missing 343 measurements, and the data needed to be further divided into 80% training and 20% test datasets. 344

The limited sample size may explain why we did not achieve high ROC-AUCs despite the inclusion of the clinical data. These challenges highlight the need for larger datasets to be collected.

In conclusion, this study presents empirical evidence of heterogeneity in lesion transition dynamics and symptom severity among mpox cases. It also suggests that the viral load in peripheral blood at lesion onset may be linked to this heterogeneity. Given the absence of currently available sensitive, simple and rapid biomarkers, the onset viral load could potentially be used as a predictor of mpox skin lesion severity. Identifying risk factors for severe symptoms at the individual level is crucial for improving treatment strategies and public health interventions against current and future outbreaks of mpox.

355 Methods

356 Ethics statement

The prospective observational study of the clinical natural history of human MPXV 357 infections at the remote Kole Hospital in the rainforest of the Congo River basin of the DRC, which 358 was conducted from March 2007 until August 2011, was reviewed and approved by the Human 359 360 Use Committee of the United States Army Medical Research Institute of Infectious Diseases (FY05-13) and the Headquarters, United States Army Medical Research and Development 361 Command Institutional Review Board (IRB), Frederick, MD, USA, as well as the Ethics Committee 362 at the University of Kinshasa School of Public Health (KSPH), Kinshasa, DRC (see [27] for details). 363 The present study was approved by the ethics committee of Nagoya University (approval number: 364 365 Hc 22-07).

366

367 Study data

368 The count of lesions on nine distinct skin regions and the oropharynx of each patient's body (a total of 10 areas) was documented and reported in [27]. In this study, we used the total 369 number of lesions for each stage (macules, papules, vesicles, pustules, umbilication, scabbing, 370 and desquamation) by aggregating the lesion count from all areas for each day of the study. 371 Additionally, we incorporated viral load data, measuring the amount of MPXV genomic DNA in 372 373 peripheral blood, pharyngeal swabs, and scabs. Clinical laboratory data, obtained through 374 hematology analysis, urinalysis, and clinical chemistry from the same group of patients, was also 375 considered. Detailed information on data preparation can be found in [27].

376

377 Modeling lesion transition dynamics

In our recent report [27], we categorized lesions into seven different stages: macule, papule, vesicle, pustule, umbilicated, scab, and desquamation. Because of the limited number of macule, papule, and vesicle lesions for the purpose of parameter estimation and the lack of clinical relevance in distinguishing between these three stages, we merged these into one stage.

The notation $L_1(t)$ therefore represents the total number of macule, papule, and vesicle lesions at time *t* since lesion onset (i.e., t = 0 represents the day at which a patient reported its first lesion appearance). Additionally, considering that the infectious period persists until all scabs or crusts have fallen off [28], we considered lesion counts in the umbilicated, pustule, and scab stages as $L_2(t)$, $L_3(t)$, and $L_4(t)$, respectively, while excluding lesion data from the desquamation stage. Subsequently, we developed the following mathematical model to describe the dynamics of lesion transitions in MPXV-infected patients:

389
$$\frac{dL_1(t)}{dt} = ge^{-ct}L_1(t) - a_1(1 - e^{-\tau_1 t})L_1(t), \tag{1}$$

390
$$\frac{dL_2(t)}{dt} = a_1(1 - e^{-\tau_1 t})L_1(t) - a_2(1 - e^{-\tau_2 t})L_2(t), \qquad (2)$$

391
$$\frac{dL_3(t)}{dt} = a_2(1 - e^{-\tau_2 t})L_2(t) - a_3(1 - e^{-\tau_3 t})L_3(t), \qquad (3)$$

392
$$\frac{dL_4(t)}{dt} = a_3(1 - e^{-\tau_3 t})L_3(t) - dL_4(t).$$
(4)

Here, we consider that the number of macule, papule, and vesicle lesions (i.e., L_1) increases at 393 per lesion rate ge^{-ct} as the infection progresses, where g corresponds to the maximum growth 394 395 rate of the number of lesions and *c* determines the rate at which this growth rate decreases over 396 time due to viral clearance. It is assumed that the lesion counts in L_1 start to exponentially 397 increase after t_0 days post first lesion onset and L_1 remains at 1 until that time (i.e., $L_1(t) = 1$ 398 and $L_2(t) = L_3(t) = L_4(t) = 0$ for $0 < t \le t_0$. The differential equations above therefore apply when the time since lesion onset $t > t_0$. In those equations, the number of scab lesions (i.e., L_4) 399 decreases at per lesion rate d, and the transition rates between the stages are denoted by 400 $a_n(1-e^{-\tau_n t})$ for n=1,2,3 where a_n is the maximum transition rate and $1-e^{-\tau_n t}$ sets the 401 time dependence of delays between the stage transitions. We estimated the above parameters, 402 i.e., g, c, a_n , d, t_0 , τ_n for n = 1,2,3, by model fitting (see below). 403

404

405 Modeling decay dynamics of viral load

(5)

406 To describe MPXV dynamics in peripheral blood, we used a simple decay model, which 407 was previously used in a study of mpox [39]:

408
$$\frac{dV(t)}{dt} = -\delta V(t),$$

where the variable V(t) is the viral RNA load (copies/mL) at time t and parameter δ represents the viral clearance rate. Note that the variable t again represents time after lesion onset; t = 0is thus the date on which the first lesion was identified.

412

413 **Parameter estimation**

We estimated the parameters of both the lesion transition and viral load models using a 414 nonlinear mixed-effect modeling approach, which incorporates fixed effects as well as random 415 effects describing the inter-patient variability in parameter values. Each parameter of patient k, 416 $\theta_k (= \theta \times e^{\pi_k})$ is decomposed as product of θ (the fixed effect) and e^{π_k} (the random effect), 417 where π_k is assumed to be drawn from a normal distribution: $N(0, \Omega)$. The fixed effect 418 parameters and the standard deviations of random effects, Ω , were estimated using the 419 stochastic approximation expectation/maximization (SAEM) algorithm, and individual parameters 420 for each patient were then estimated using the empirical Bayes method. A right-truncated normal 421 distribution was used in the likelihood function to account for the left censoring of the lesion count 422 423 data or viral load data (i.e., when the lesion count or viral load is not detectable) [59]. MONOLIX 2021R2 (www.lixoft.com), a program for maximum likelihood estimation for nonlinear mixed-effect 424 models, was used to fit the model to these data. We changed the initial values multiple times to 425 426 avoid obtaining a local minimum of the negative log-likelihood function and to confirm the 427 robustness of our parameter estimates.

428

429 Unsupervised clustering and stratification of lesion transition dynamics

430 We extracted three key "features" of lesion transition dynamics for each individual: i) the 431 average macule, papule, and vesicle lesion count increase rate (i.e., $\frac{1}{T} \int_{t_0}^{T+t_0} g e^{-ct} dt$) over 40

days (i.e., T = 40); ii) the average stage transition rates (i.e., $\frac{1}{T} \int_{t_0}^{T+t_0} a_n (1 - e^{-\tau_n t}) dt$ for n =432 1,2,3), again with T = 40; and iii) the scab lesion decrease rate, d. For the purpose of 433 434 stratification of the lesion transition dynamics, unsupervised random forest clustering based on each of five features was used (rfUtilities in R). The use of random forest allowed us to avoid 435 overfitting by bootstrap aggregating (bagging) and to achieve better generalization performance 436 437 [60]. Specifically, after a random forest dissimilarity (i.e., the distance matrix between all pairs of 438 samples) was obtained, it was visualized by using UMAP in a two-dimensional plane and was stratified with spectral clustering (Python scikit-learn). The optimal number of clusters was 439 440 determined using the eigengap heuristic method [61].

441

442 **Classifying the stratified groups with a logistic regression**

We built a predictive model for the stratified groups using a logistic regression. In this model, we used the viral load at lesion onset as the predictor variable, with the classification of the stratified group as the outcome variable. To ensure that the model was both robust and generalizable, we subjected it to rigorous validation through a 6-fold cross-validation process. We assessed the model performance using AUC scores and reported the mean and standard deviation of these scores obtained from 50 replications [46].

449

450 **Classifying the stratified groups with a LightGBM model**

Since some features showed high correlation or had missing values, we used the gradient boosting machine learning algorithm, specifically LightGBM, as our primary method of analysis [62]. LightGBM was chosen for its decision tree-based algorithm, which inherently handles the multicollinearity of predictors effectively. Given the limited number of data samples, we did not fine-tune the hyperparameters of LightGBM and the model was constructed using its default settings.

457

458 Data preparation

- To handle missing values, we used KNNImputer [63], employing a k-nearest neighbor approach to impute missing values based on the mean of their nearest neighbors. During this process, missing values were addressed across four neighborhoods.
- 462

463 **Data splitting and training algorithm**

The stratification groups consisted of 42, 76, and 31 patients for G1, G2, and G3, respectively. 464 We approached the problem as a binary classification task, where the model aimed to predict 465 466 whether a given individual belonged to G1 or not (similarly for G2 and G3). The model then assigned each patient to their most likely group. We split the stratified groups into two datasets: 467 80% as a training dataset and 20% as a test dataset. The former was used for model training 468 and internal validation, and the latter served as a hold-out dataset for external validation. To 469 470 address class imbalance, we generated synthetic instances of the minority class in the training 471 data using the Synthetic Minority Over-sampling Technique (SMOTE) [64].

472

473 Feature selection and model evaluation

474 Because of the propensity of machine learning models to overfit when the data contain a high 475 number of features, feature selection was necessary. Feature selection hinged on the significance of SHAP values; we incrementally added features to the machine learning model 476 477 according to their importance and selected those that yielded the highest ROC-AUC in the training dataset. The process was repeated 50 times with different pseudorandom number 478 479 generator initiation seeds to account for the influence of differences in splits. Our model performance was evaluated using AUC scores, and we reported the mean and standard 480 deviation of these scores obtained from 50 replications. 481

482

483 Statistical analysis

484 When necessary, variables were compared among different groups using analysis of 485 variance (the Kruskal-Wallis test and Fisher's exact test, for more than two numerical or categorial 486 variables, respectively) with false discovery rate, or the Mann-Whitney U test (for two numerical variables) with Bonferroni correction. All statistical analyses were performed using R (version 487 488 4.2.0).

LIST OF SUPPLEMENTARY MATERIALS 489

- Supplementary Figure 1 | Reconstructed lesion transition dynamics by stratified group for 149 490
- individuals 491
- Supplementary Table 1 | The estimated fixed and random effects for lesion transition dynamics 492

medRxiv preprint doi: https://doi.org/10.1101/2024.01.28.24301907; this version posted January 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in

perpetuity. It is made available under a CC-BY-ND 4.0 International license .

REFERENCES 493

- 494 Guarner, J., C. Del Rio, and P.N. Malani, Monkeypox in 2022-What Clinicians Need to Know. 1. 495 JAMA, 2022. 328(2): p. 139-140.
- Ladnyj, I.D., P. Ziegler, and E. Kima, A human infection caused by monkeypox virus in 496 2. 497 Basankusu Territory, Democratic Republic of the Congo. Bull World Health Organ, 1972. 46(5): 498 p. 593-7.
- 499 3. Catala, A., et al., Monkeypox outbreak in Spain: clinical and epidemiological findings in a 500 prospective cross-sectional study of 185 cases. Br J Dermatol, 2022. 187(5): p. 765-772.
- Breman, J.G., et al., Human monkeypox, 1970-79. Bull World Health Organ, 1980. 58(2): p. 165-501 4. 502 82.
- Heymann, D.L., M. Szczeniowski, and K. Esteves, Re-emergence of monkeypox in Africa: a 503 5. 504 review of the past six years. Br Med Bull, 1998. 54(3): p. 693-702.
- 505 6. Mandja, B.M., et al., Temporal and Spatial Dynamics of Monkeypox in Democratic Republic of 506 Congo, 2000-2015. Ecohealth, 2019. 16(3): p. 476-487.
- Nguyen, P.Y., et al., Reemergence of Human Monkeypox and Declining Population Immunity in 507 7. 508 the Context of Urbanization, Nigeria, 2017-2020. Emerg Infect Dis, 2021. 27(4): p. 1007-14.
- 509 Ogoina, D., et al., Clinical Course and Outcome of Human Monkeypox in Nigeria, Clin Infect Dis, 8. 510 2020. **71**(8): p. e210-e214.
- 511 9. Isidro, J., et al., Phylogenomic characterization and signs of microevolution in the 2022 multi-512 country outbreak of monkeypox virus. Nat Med, 2022. 28(8): p. 1569-1572.
- 513 Thornhill, J.P., et al., Monkeypox Virus Infection in Humans across 16 Countries - April-June 10. 514 2022. N Engl J Med, 2022. 387(8): p. 679-691.
- WHO. 2022-23 Mpox (Monkeypox) Outbreak: Global Trends. [cited 2023 December, 26]; 515 11. 516 Available from: https://worldhealthorg.shinyapps.io/mpx_global/.
- 517 12. Zhu, J., et al., *Exploring the key genomic variation in monkeypox virus during the 2022 outbreak*. 518 BMC Genom Data, 2023. 24(1): p. 67.
- WHO. *Mpox (monkeypox) outbreak*. [cited 2023 December, 26]; Available from: 519 13. 520 https://www.who.int/europe/emergencies/situations/monkeypox.
- 521 14. Akhmetzhanov, A.R., L. Ponce, and R.N. Thompson, *Emergence potential of monkeypox in the* 522 Western Pacific Region, July 2022. Int J Infect Dis, 2022. 122: p. 829-831.
- 523 Endo, A., S.M. Jung, and F. Miura, Mpox emergence in Japan: ongoing risk of establishment in 15. 524 Asia. Lancet, 2023. 401(10392): p. 1923-1924.
- Kibungu, E.M., et al., Clade I-Associated Mpox Cases Associated with Sexual Contact, the 525 16. 526 Democratic Republic of the Congo. Emerg Infect Dis, 2024. 30(1): p. 172-176.
- 527 17. Jezek, Z., et al., Human monkeypox: clinical features of 282 patients. J Infect Dis, 1987. 156(2): 528 p. 293-8.
- 529 Americo, J.L., P.L. Earl, and B. Moss, Virulence differences of mpox (monkeypox) virus clades I, 18. 530 IIa, and IIb.1 in a small animal model. Proc Natl Acad Sci U S A, 2023. 120(8): p. e2220415120.
- 531 19. Beer, E.M. and V.B. Rao, A systematic review of the epidemiology of human monkeypox 532 outbreaks and implications for outbreak strategy. PLoS Negl Trop Dis, 2019. 13(10): p. 533 e0007791.
- 534 20. Chen, N., et al., Virulence differences between monkeypox virus isolates from West Africa and the Congo basin. Virology, 2005. 340(1): p. 46-63. 535
- 536 21. Hudson, P.N., et al., Elucidating the role of the complement control protein in monkeypox 537 pathogenicity. PLoS One, 2012. 7(4): p. e35086.
- 538 Yinka-Ogunleye, A., et al., Outbreak of human monkeypox in Nigeria in 2017-18: a clinical and 22. 539 epidemiological report. Lancet Infect Dis, 2019. 19(8): p. 872-879.
- 540 23. WHO. Mpox (monkeypox)- Democratic Republic of the Congo. [cited 2023 December, 26]; 541 Available from: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON493.
- 542 Adler, H., et al., *Clinical features and management of human monkeypox: a retrospective* 24. 543 observational study in the UK. Lancet Infect Dis, 2022. 22(8): p. 1153-1162.

medRxiv preprint doi: https://doi.org/10.1101/2024.01.28.24301907; this version posted January 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in

perpetuity. It is made available under a CC-BY-ND 4.0 International license .

544 25. Huhn, G.D., et al., Clinical characteristics of human monkeypox, and risk factors for severe 545 disease. Clin Infect Dis, 2005. 41(12): p. 1742-51. 546 26. Patel, A., et al., Clinical features and novel presentations of human monkeypox in a central 547 London centre during the 2022 outbreak: descriptive case series. BMJ, 2022. 378: p. e072410. 548 Pittman, P.R., et al., Clinical characterization and placental pathology of mpox infection in 27. 549 hospitalized patients in the Democratic Republic of the Congo. PLoS Negl Trop Dis, 2023. 17(4): 550 p. e0010384. 551 28. CDC. Clinical Recognition. Available from: 552 https://www.cdc.gov/poxvirus/mpox/clinicians/clinical-recognition.html. 553 29. Di Giulio, D.B. and P.B. Eckburg, Human monkeypox: an emerging zoonosis. Lancet Infect Dis, 554 2004. 4(1): p. 15-25. 555 30. Titanji, B.K., et al., Monkeypox: A Contemporary Review for Healthcare Professionals. Open Forum Infect Dis, 2022. 9(7): p. ofac310. 556 557 31. CDC. Isolation and Infection Control At Home. [cited 2023 December, 26]; Available from: 558 https://www.cdc.gov/poxvirus/mpox/clinicians/infection-control-home.html. 559 32. Prasad, S., et al., A dermatologic assessment of 101 mpox (monkeypox) cases from 13 countries 560 during the 2022 outbreak: Skin lesion morphology, clinical course, and scarring. J Am Acad 561 Dermatol, 2023. 88(5): p. 1066-1073. 562 33. Suner, C., et al., Viral dynamics in patients with monkeypox infection: a prospective cohort study 563 in Spain. Lancet Infect Dis, 2023. 23(4): p. 445-453. 564 34. Tarin-Vicente, E.J., et al., Clinical presentation and virological assessment of confirmed human 565 monkeypox virus cases in Spain: a prospective observational cohort study. Lancet, 2022. 566 **400**(10353): p. 661-669. 567 Li, P., et al., Clinical Features, Antiviral Treatment, and Patient Outcomes: A Systematic Review 35. 568 and Comparative Analysis of the Previous and the 2022 Mpox Outbreaks. J Infect Dis, 2023. 569 228(4): p. 391-401. CDC. Treatment Information for Healthcare Professionals. [cited 2023 December, 26]; 570 36. 571 Available from: https://www.cdc.gov/poxvirus/mpox/clinicians/treatment.html. 572 37. Kawakami, E., et al., Monitoring of blood biochemical markers for periprosthetic joint infection 573 using ensemble machine learning and UMAP embedding. Arch Orthop Trauma Surg, 2023. 574 **143**(10): p. 6057-6067. 575 38. Shi, T. and S. Horvath, Unsupervised learning with random forest predictors. Journal of 576 Computational and Graphical Statistics, 2006. 15(1): p. 118-138. Jeong, Y.D., et al., *Modelling the effectiveness of isolation strategies for managing mpox* 577 39. 578 outbreaks with variable infectiousness profiles. medRxiv, 2023: p. 2023.10.04.23296551. 579 40. Hazra, A., et al., Mpox in people with past infection or a complete vaccination course: a global 580 case series. Lancet Infect Dis, 2024. 24(1): p. 57-64. 581 Jiang, R.M., et al., Diagnosis, treatment, and prevention of monkeypox in children: an experts' 41. 582 consensus statement. World J Pediatr, 2023. 19(3): p. 231-242. McNeil, A.J., et al., Counting Monkeypox Lesions in Patient Photographs: Limits of Agreement of 583 42. 584 Manual Counts and Artificial Intelligence. J Invest Dermatol, 2023. 143(2): p. 347-350 e4. 585 43. Jones, B., et al., Variability in clinical assessment of clade IIb mpox lesions. Int J Infect Dis, 2023. 586 **137**: p. 60-62. 587 Lapa, D., et al., *Monkeypox virus isolation from a semen sample collected in the early phase of* 44. 588 infection in a patient with prolonged seminal viral shedding. Lancet Infect Dis, 2022. 22(9): p. 589 1267-1269. 590 Lundberg, S.M. and S.I. Lee, A Unified Approach to Interpreting Model Predictions. Advances in 45. 591 Neural Information Processing Systems 30 (Nips 2017), 2017. 30. 592 46. Bluthgen, C., et al., Computed tomography radiomics for the prediction of thymic epithelial tumor 593 histology, TNM stage and myasthenia gravis. PLoS One, 2021. 16(12): p. e0261401. 594 47. Silva, S., et al., Clinical and laboratory diagnosis of monkeypox (mpox): Current status and future 595 directions. iScience, 2023. 26(6): p. 106759.

medRxiv preprint doi: https://doi.org/10.1101/2024.01.28.24301907; this version posted January 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

perpetuity. It is made available under a CC-BY-ND 4.0 International license .

- Wang, Z., et al., *The human host response to monkeypox infection: a proteomic case series study*.
 EMBO Mol Med, 2022. 14(11): p. e16643.
- 49. Rizk, J.G., et al., *Prevention and Treatment of Monkeypox*. Drugs, 2022. **82**(9): p. 957-963.
- 599 50. Siegrist, E.A. and J. Sassine, *Antivirals With Activity Against Mpox: A Clinically Oriented* 600 *Review.* Clin Infect Dis, 2023. **76**(1): p. 155-164.
- 51. van Ewijk, C.E., et al., Mpox outbreak in the Netherlands, 2022: public health response,
 characteristics of the first 1,000 cases and protection of the first-generation smallpox vaccine.
 Euro Surveill, 2023. 28(12).
- 604 52. Reynolds, M.G., et al., *Clinical manifestations of human monkeypox influenced by route of infection*. J Infect Dis, 2006. **194**(6): p. 773-80.
- 606 53. Mitja, O., et al., *Monkeypox*. Lancet, 2023. **401**(10370): p. 60-74.
- 60754.WHO. Multi-country outbreak of mpox, External situation report#31 22 December 2023.608[cited 2023 December, 26]; Available from: https://www.who.int/publications/m/item/multi-country-outbreak-of-mpox--external-situation-report-31---22-december-2023.
- 55. Karagoz, A., et al., Monkeypox (mpox) virus: Classification, origin, transmission, genome
 organization, antiviral drugs, and molecular diagnosis. J Infect Public Health, 2023. 16(4): p.
 531-541.
- 613 56. Sherwat, A., et al., *Tecovirimat and the Treatment of Monkeypox Past, Present, and Future*614 *Considerations.* N Engl J Med, 2022. 387(7): p. 579-581.
- 615 57. Akiyama, Y., et al., *Efficacy and viral dynamics of tecovirimat in patients with MPOX: A*616 *multicenter open-label, double-arm trial in Japan.* J Infect Chemother, 2023.
- 617 58. Matias, W.R., et al., *Tecovirimat for the Treatment of Human Monkeypox: An Initial Series From*618 *Massachusetts, United States.* Open Forum Infect Dis, 2022. 9(8): p. ofac377.
- 59. Samson, A., M. Lavielle, and F. Mentré, *Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: Application to HIV dynamics model.* Computational Statistics & Data Analysis, 2006. 51(3): p. 1562-1574.
- 60. Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H., *The elements of statistical learning: data mining, inference, and prediction.* Vol. 2. 2009: Springer.
- 624 61. von Luxburg, U., *A tutorial on spectral clustering*. Statistics and Computing, 2007. 17(4): p. 395625 416.
- 626 62. Ke, G.a.M., Qi and Finley, Thomas and Wang, Taifeng and Chen, Wei and Ma, Weidong and Ye,
 627 Qiwei and Liu, Tie-Yan, *Lightgbm: A highly efficient gradient boosting decision tree*. Advances
 628 in neural information processing systems, 2017. 30: p. 3146-3154.
- 629 63. Troyanskaya, O., et al., *Missing value estimation methods for DNA microarrays*. Bioinformatics,
 630 2001. 17(6): p. 520-5.
- 631 64. Chawla, N.V., et al., *SMOTE: Synthetic minority over-sampling technique*. Journal of Artificial
 632 Intelligence Research, 2002. 16: p. 321-357.

633 ACKNOWLEDGMENTS

634 This study was supported in part by Grant-in-Aid for JSPS Fellows 23KJ1081 (to T.N.); Scientific Research (KAKENHI) B 23H03497 (to S.I.): Grant-in-Aid for Transformative Research 635 Areas 22H05215 (to S.I.); Grant-in-Aid for Challenging Research (Exploratory) 22K19829 (to S.I.); 636 AMED CREST 19gm1310002 (to S.I.); AMED Research Program on Emerging and Re-emerging 637 Infectious Diseases 22fk0108509 (to S.I.), 23fk0108684 (to S.I.), 23fk0108685 (to S.I.); AMED 638 Research Program on HIV/AIDS 22fk0410052 (to S.I.); AMED Program for Basic and Clinical 639 Research on Hepatitis 22fk0210094 (to S.I.); AMED Program on the Innovative Development and 640 the Application of New Drugs for Hepatitis B 22fk0310504h0501 (to S.I.); AMED Strategic 641 Research Program for Brain Sciences 22wm0425011s0302; AMED JP22dm0307009 (to K.A.); 642 643 JST MIRAI JPMJMI22G1 (to S.I.); Moonshot R&D JPMJMS2021 (to K.A. and S.I.) and 644 JPMJMS2025 (to S.I.); Institute of AI and Beyond at the University of Tokyo (to K.A.); Shin-Nihon of Advanced Medical Research (to S.I.); SECOM Science and Technology Foundation (to S.I.); 645 646 The Japan Prize Foundation (to S.I.). The collaboration between R.N.T. and S.I. was supported by a Royal Society International Exchange award (grant number IES-R3-193037). 647

648

649 **AUTHOR CONTRIBUTIONS**

650 SI and FM designed the research. TN carried out the computational analysis. SI, FM, 651 JWH, and PRP supervised the project. All authors discussed the research and contributed to 652 writing the manuscript.

653

654 COMPETING FINANCIAL INTERESTS

- The authors declare no conflicts of interest associated with this manuscript.
- 656

657 INSTITUTIONAL REVIEW BOARD STATEMENT

This study was approved by the ethics committees of Nagoya University (hc22-05).



661 Extended Data Figure 1 | Flowchart of the cohort, along with the number of participants

662 and inclusion criteria for our analysis.

medRxiv preprint doi: https://doi.org/10.1101/2024.01.28.24301907; this version posted January 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.



663

Extended Data Figure 2 | Time course distribution of the average lesion count of each
 stage: Time course plots of individual lesion counts for each stage (i.e., macule, papule, vesicle,
 pustule, umbilicated, scab, and desquamation) across all patients.

Onset of lesion (Days)



668 The lesion counts in the macule, papule, and vesicle (i.e., these three are in one compartment, 669 as described in the main text); pustule; umbilicated; and scab stages are defined as L_1 , L_2 , L_3 , 670 and L_4 , respectively. It is assumed that L_1 increases after t_0 days post lesion onset according 671 to a per-lesion rate ge^{-ct} as the infection progresses (with $L_1(t) = 1$ for $t \le t_0$), where g 672 corresponds to the maximum growth rate of the number of lesions and *c* determines the rate at 673 which this growth rate decreases over time. In contrast, L_4 decreases according to a rate d, and 674 the transition rates between the stages are denoted as $a_n(1-e^{-\tau_n t})$ for n = 1,2,3 where a_n is 675 the maximum transition rate and $1 - e^{-\tau_n t}$ sets the time dependence of the delay between the 676 stage transitions. 677



679 Extended Data Figure 4 | The time-course patterns of reconstructed lesion dynamics for 680 each stage: The reconstructed individual lesion dynamics for each stage, shown for each

stratified group (i.e., G1, G2, and G3).





683 **Extended Data Figure 5 | Relationship between lesion dynamics and virus dynamics:** 684 Correlations between the features of individual-level lesion transition dynamics (i.e., the lesion 685 duration and total lesion counts) and those of individual-level decay viral load dynamics (i.e., the 686 viral clearance rate, the total viral load and the viral load at lesion disappearance).



687

Extended Data Figure 6 | Distribution of features selected by machine learning: (A) and (B) show distributions of selected features by LightGBM with G1 or not, and G3 or not, respectively. Statistical significance is calculated using the pairwise Mann-Whitney U test. (NS.: p-value > 0.05, :: p-value ≤ 0.05 , :: p-value ≤ 0.01 , and :: p-value ≤ 0.001 , respectively). The statistical significance of the difference is indicated on the top of the plots.

Supplementary Information

Modeling lesion transition dynamics to clinically characterize mpox patients in the Democratic Republic of the Congo

Takara Nishiyama^{1,†}, Fuminari Miura^{2,3,†}, Yong Dam Jeong¹, Naotoshi Nakamura¹, Masahiro Ishikane⁴, Noriko Iwamoto⁴, Michiyo Suzuki⁴, Ayana Sakurai⁴, Kazuyuki Aihara⁵, Koichi Watashi⁶, William S Hart⁷, Robin N Thompson⁷, Yasuhiro Yasutomi^{8,9}, Norio Ohmagari⁴, Placide Mbala Kingebeni¹⁰, John W. Huggins¹¹, Shingo Iwami^{1,5,12,13,14,15,16,‡,*}, Phillip R. Pittman^{9,‡,*}

¹interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan.²Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands. ³Center for Marine Environmental Studies (CMES), Ehime University, Ehime, Japan. ⁴Disease Control and Prevention Centre, National Centre for Global Health and Medicine, Tokyo, Japan. ⁵International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo, Japan. ⁶Research Center for Drug and Vaccine Development, National Institute of Infectious Diseases, Tokyo, Japan. ⁷Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK. ⁸Tsukuba Primate Research Center, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Tsukuba, Japan. ⁹School of Integrative and Global Majors, University of Tsukuba, Tsukuba, Japan. ¹⁰Institut National de Recherche Biomédicale, Ministère de la Santé Publique, Democratic Republic of the Congo, ¹¹Division of Medicine, U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID), Fort Detrick, Maryland, United States of America. ¹²Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan.¹³Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Kyoto, Japan.¹⁴Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Saitama, Japan.¹⁵NEXT-Ganken Program, Japanese Foundation for Cancer Research (JFCR), Tokyo, Japan. ¹⁶Science Groove Inc., Fukuoka, Japan.

G1



G2 ID-7 4 uno units of 0, 0 4 0 ID-24 Uno 3 UD-24 Uno 3 Uno 24 Uno 24 Uno 24 Uno 24 Uno 24 Uno 24 Uno 3 Uno 24 Un ID-22 44 3-10-22 10-00 10-01 10-01 ID- 6 ID-26 III 03-ID- 8 ID- 27 4 3 2 0 1010 10 0 10 0 0 10 20 30 40 $u_{\text{diff}}^{\text{diff}} = \frac{1}{10^{-10}} \frac{1}{10^{-$ ID- 60 ID- 84 4 1010 resieu contra 2 1010 1 1 Dg10 Lesion count 0 10 20 30 40 Onset of lesion (Days) ID- 140 ID- 153 og 10 Lesion count Other of immunous of the second secon $\begin{array}{c} \text{Other D Head (LDHy)} & \text{Other D Head$ 1 30 n (Days) 10-200 4 4 3 4 4 3 4 4 3 4 4 3 4 4 4 3 4 0 0 10 20 50 40 0 0 met of teston (Days) ID- 219 1 N Solution count ID- 225 10- 255 10- 255 10-ID- 226 ID- 230 ID- 239 ID- 240 4 - Contraction Countraction Co ID- 250 Log10 Lesion count - 210 Lesion count 0 10 20 30 40 Onset of lacion (Dava)



Supplementary Figure 1 | Reconstructed lesion transition dynamics for stratified groups of 149 individuals: Individual-level model fits of lesion count data using the model described in Eqs. (1-4) in the main text. The closed dots and solid curves indicate measured data and estimated lesion dynamics, respectively. Dots and curves are colored according to stages (i.e., L_1 , L_2 , L_3 , and L_4 : orange, magenta, light purple, and cyan).

Supplementary Table 1 | Estimated fixed and random effects for lesion transition dynamics

Parameters	Description	Unit	artheta: Fixed effect (SE)*	Ω : SD of random effect (SE)*
g	Maximum growth rate of the lesion	1/day	2.58 (0.07)	0.18 (0.02)
С	Decreasing rate of maximum growth rate of the lesion	1/day	0.53 (0.03)	0.48 (0.04)
a_1	Maximum transition rate from L_1 to L_2	1/day	1.80 (0.11)	0.48 (0.05)
<i>a</i> ₂	Maximum transition rate from L_2 to L_3	1/day	1.63 (0.17)	0.82 (0.07)
<i>a</i> ₃	Maximum transition rate from L_3 to L_4	1/day	1.97 (0.17)	0.68 (0.06)
d	Decreasing rate of L_4	1/day	0.26 (0.01)	0.38 (0.04)
t_0	Day at increasing lesion count	Days	0.53 (0.03)	0.31 (0.03)
$ au_1$	Scaling coefficient for transition delay from L_1 to L_2	1/day	$9.9 \times 10^{-3} (0.22 \times 10^{-3})$	0.05 (0.02)
$ au_2$	Scaling coefficient for transition delay from L_2 to L_3	1/day	$9.3 \times 10^{-2} (1.83 \times 10^{-2})$	1.56 (0.16)
$ au_3$	Scaling coefficient for transition delay from L_3 to L_4	1/day	$4.7 \times 10^{-2} (0.38 \times 10^{-2})$	0.40 (0.05)

* The parameter for patient k, $\vartheta_i (= \vartheta \times e^{\pi_k})$ is represented as a product of ϑ (a fixed effect) and e^{π_k} (a random effect). π_k follows the normal distribution with mean 0 and standard deviation Ω . SE: standard error.