THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Somatic evolution and global expansion of an ancient transmissible cancer lineage

Citation for published version:
Baez-Ortega, A, Gori, K, Strakova, A, Allen, JL, Allum, KM, Bansse-Issa, L, Bhutia, TN, Bisson, JL, Briceño, C, Castillo Domracheva, A, Corrigan, AM, Cran, HR, Crawford, JT, Davis, E, de Castro, KF, B de Nardi, A, de Vos, AP, Delgadillo Keenan, L, Donelan, EM, Espinoza Huerta, AR, Faramade, IA, Fazil, M, Fotopoulou, E, Fruean, SN, Gallardo-Arrieta, F, Glebova, O, Gouletsou, PG, Häfelin Manrique, RF, Henriques, JJGP, Horta, RS, Ignatenko, N, Kane, Y, King, C, Koenig, D, Krupa, A, Kruzeniski, SJ, Kwon, Y-M, Lanza-Perea, M, Lazyan, M, Lopez Quintana, AM, Losfelt, T, Marino, G, Martínez Castañeda, S, Martínez-López, MF, Meyer, M, Migneco, EJ, Nakanwagi, B, Neal, KB, Neunzig, W, Ní Leathlobhair, M, Nixon, SJ, Ortega-Pacheco, A, Pedraza-Ordoñez, F, Peleteiro, MC, Polak, K, Pye, RJ, Reece, JF, Rojas Gutierrez, J, Sadia, H, Schmeling, SK, Shamanova, O, Sherlock, AG, Stammnitz, M, Steenland-Smit, AE, Svitich, A, Tapia Martínez, LJ, Thoya Ngoka, I, Torres, CG, Tudor, EM, van der Wel, MG, Vițălaru, BA, Vural, SA, Walkinton, O, Wang, J, Wehrle-Martinez, AS, Widdowson, SAE, Stratton, MR, Alexandrov, LB, Martincorena, I & Murchison, EP 2019, 'Somatic evolution and global expansion of an ancient transmissible cancer lineage', *Science*, vol. 365, no. 6452. https://doi.org/10.1126/science.aau9923

Digital Object Identifier (DOI):
10.1126/science.aau9923

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Science

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 12. Oct. 2024

Somatic evolution and global expansion of an ancient transmissible cancer lineage

Adrian Baez-Ortega[1], Kevin Gori[1*], Andrea Strakova[1*], Janice L. Allen[2], Karen M. Allum[3], Leontine Bansse-Issa[4], Thinlay N. Bhutia[5], Jocelyn L. Bisson[1], Cristóbal Briceño[6], Artemio Castillo Domracheva[7], Anne M. Corrigan[8], Hugh R. Cran[9], Jane T. Crawford[10], Eric Davis[11], Karina F. de Castro[12], Andrigo B. de Nardi[13], Anna P. de Vos[14], Laura Delgadillo Keenan[15], Edward M. Donelan[2], Adela R. Espinoza Huerta[16], Ibikunle A. Faramade[17], Mohammed Fazil[18], Eleni Fotopoulou[19], Skye N. Fruean[20], Fanny Gallardo-Arrieta[21], Olga Glebova[22], Pagona G. Gouletsou[23], Rodrigo F. Häfelin Manrique[24], Joaquim J. G. P. Henriques[25], Rodrigo S. Horta[26], Natalia Ignatenko[27], Yaghouba Kane[28], Cathy King[3], Debbie Koenig[3], Ada Krupa[29], Steven J. Kruzeniski[16], Young-Mi Kwon[1], Marta Lanza-Perea[8], Mihran Lazyan[30], Adriana M. Lopez Quintana[31], Thibault Losfelt[32], Gabriele Marino[33], Simón Martínez Castañeda[34], Mayra F. Martínez-López[35], Michael Meyer[36], Edward J. Migneco[37], Berna Nakanwagi[38], Karter B. Neal[39], Winifred Neunzig[3], Máire Ní Leathlobhair[1], Sally J. Nixon[40], Antonio Ortega-Pacheco[41], Francisco Pedraza-Ordoñez[42], Maria C. Peleteiro[43], Katherine Polak[44], Ruth J. Pye[45], John F. Reece[46], Jose Rojas Gutierrez[47], Haleema Sadia[48], Sheila K. Schmeling[49], Olga Shamanova[50], Alan G. Sherlock[45], Maximilian Stammnitz[1], Audrey E. Steenland-Smit[4], Alla Svitich[51], Lester J. Tapia Martínez[16], Ismail Thoya Ngoka[52], Cristian G. Torres[53], Elizabeth M. Tudor[54], Mirjam G. van der Wel[55], Bogdan A. Vițălaru[56], Sevil A. Vural[57], Oliver Walkinton[45], Jinhong Wang[1], Alvaro S. Wehrle-Martinez[58], Sophie A. E. Widdowson[59], Michael R. Stratton[60], Ludmil B. Alexandrov[61], Iñigo Martincorena[60], Elizabeth P. Murchison[1§]

[1] Transmissible Cancer Group, Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

[2] Animal Management in Rural and Remote Indigenous Communities (AMRRIC), Darwin, Australia

[3] World Vets, Gig Harbor, USA

[4] Animal Shelter, Stichting Dierenbescherming Suriname, Paramaribo, Suriname

[5] Sikkim Anti-Rabies and Animal Health Programme, Department of Animal Husbandry, Livestock, Fisheries and Veterinary Services, Government of Sikkim, India

[6] ConserLab, Animal Preventive Medicine Department, Faculty of Animal and Veterinary Sciences, University of Chile, Santiago, Chile

[7] Corozal Veterinary Hospital, University of Panamá, Panama City, Republic of Panama
[8] St. George's University, True Blue, Grenada
[9] The Nakuru District Veterinary Scheme Ltd, Nakuru, Kenya
[10] Animal Medical Centre, Belize City, Belize
[11] International Animal Welfare Training Institute, UC Davis School of Veterinary Medicine, Davis, USA
[12] Centro Universitário de Rio Preto (UNIRP), São José do Rio Preto, São Paulo, Brazil
[13] Department of Clinical and Veterinary Surgery, São Paulo State University (UNESP), São Paulo, Brazil
[14] Ladybrand Animal Clinic, Ladybrand, South Africa
[15] Veterinary Clinic Sr. Dog's, Guadalajara, Mexico
[16] World Vets Latin America Veterinary Training Center, Granada, Nicaragua
[17] National Veterinary Research Institute, Vom, Nigeria
[18] Animal Clinic, Mombasa, Kenya
[19] Intermunicipal Stray Animals Care Centre (DIKEPAZ), Perama, Greece
[20] Animal Protection Society of Samoa, Apia, Samoa
[21] Faculty of Veterinary Science, University of Zulia, Maracaibo, Venezuela
[22] Veterinary Clinic BIOCONTROL, Moscow, Russia
[23] Faculty of Veterinary Medicine, School of Health Sciences, University of Thessaly, Karditsa, Greece
[24] Veterinary Clinic El Roble, Animal Healthcare Network, Faculty of Animal and Veterinary Sciences, University of Chile, Santiago de Chile, Chile
[25] OnevetGroup, Hospital Veterinário Berna, Lisboa, Portugal
[26] Universidade Vila Velha, Vila Velha, Brazil
[27] Veterinary Clinic Zoovetservis, Kiev, Ukraine
[28] École Inter-états des Sciences et Médecine Vétérinaires de Dakar, Dakar, Senegal
[29] Department of Small Animal Medicine, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands
[30] Vetexpert Veterinary Group, Yerevan, Armenia
[31] Veterinary Clinic Lopez Quintana, Maldonado, Uruguay
[32] Clinique Veterinaire de Grand Fond, Saint Gilles les Bains, Reunion, France
[33] Department of Veterinary Sciences, University of Messina, Messina, Italy
[34] Facultad de Medicina Veterinaria y Zootecnia, Universidad Autónoma del Estado de México, Toluca, Mexico
[35] School of Veterinary Medicine, Universidad de las Américas, Quito, Ecuador
[36] Touray & Meyer Vet Clinic, Serrekunda, The Gambia
[37] Hillside Animal Hospital, St. Louis, USA

[38] The Kampala Veterinary Surgery, Kampala, Uganda
[39] Asavet Veterinary Charities, Tucson, United States
[40] Vets Beyond Borders, Bylakuppe, India
[41] Faculty of Veterinary Medicine, Autonomous University of Yucatan, Merida, Mexico
[42] Laboratorio de Patología Veterinaria, Universidad de Caldas, Manizales, Colombia
[43] Interdisciplinary Centre of Research in Animal Health (CIISA), Faculty of Veterinary Medicine, University of Lisbon, Lisboa, Portugal
[44] Four Paws International, Vienna, Austria
[45] Vets Beyond Borders, The Rocks, Australia
[46] Help in Suffering, Jaipur, India
[47] Veterinary Clinic Dr José Rojas, Los Andes, Chile
[48] Department of Biotechnology, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan
[49] Corozal Veterinary Clinic, Corozal Town, Belize
[50] Veterinary Clinic Vetmaster, Ramenskoye, Russia
[51] State Hospital of Veterinary Medicine, Dniprodzerzhynsk, Ukraine
[52] Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya
[53] Laboratory of Biomedicine and Regenerative Medicine, Department of Clinical Sciences, Faculty of Animal and Veterinary Sciences, University of Chile, Santiago, Chile
[54] Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Melbourne, Australia
[55] Animal Anti Cruelty League, Port Elizabeth, South Africa
[56] Clinical Sciences Department, Faculty of Veterinary Medicine Bucharest, Bucharest, Romania
[57] Department of Pathology, Faculty of Veterinary Medicine, Ankara University, Ankara, Turkey
[58] Faculty of Veterinary Sciences, National University of Asuncion, San Lorenzo, Paraguay
[59] Lilongwe Society for Protection and Care of Animals (LSPCA), Lilongwe, Malawi
[60] Wellcome Sanger Institute, Hinxton, UK
[61] Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, USA
* These authors contributed equally
[§]Corresponding author, epm27@cam.ac.uk

Structured Abstract

INTRODUCTION

The canine transmissible venereal tumour (CTVT) is a sexually transmitted cancer that manifests as genital tumours in dogs. This cancer first arose in an individual 'founder dog' several thousand years ago, and has since survived by transfer of living cancer cells to new hosts during coitus. Today, CTVT affects dogs around the world and is the oldest and most prolific known cancer lineage. CTVT thus provides a unique opportunity to explore the evolution of cancer over the long-term, and to track the unusual biological transition from multicellular organism to obligate conspecific asexual parasite. Furthermore, the CTVT genome, acting as a living biomarker, has recorded the changing mutagenic environments experienced by this cancer throughout millennia and across continents.

RATIONALE

To capture the genetic diversity of the CTVT lineage, we analysed somatic mutations extracted from the protein-coding genomes (exomes) of 546 globally distributed CTVT tumours. We inferred a time-resolved phylogenetic tree for the clone and used this to trace the worldwide spread of the disease and to select subsets of mutations acquired at known geographical locations and time-periods. Computational methods were applied to extract mutational signatures and to measure their exposures across time and space. In addition, we assessed the activity of selection using ratios of non-synonymous and synonymous variants.

RESULTS

The CTVT phylogeny reveals that the lineage first arose from its founder dog 4,000–8,500 years ago, likely in Asia, with the most recent common ancestor of modern globally distributed tumours occurring ~1,900 years ago. CTVT underwent a rapid global expansion within the last 500 years, likely aided by intensification of human maritime travel. We identify a highly specific mutational signature dominated by C>T mutations at GTCCA pentanucleotide contexts which operated in CTVT up until ~1,000 years ago. The number of mutations caused by ultraviolet light exposure is correlated with latitude of tumour collection, and we identify CTVTs with heritable hyperactivity of an endogenous mutational process. Several 'driver' mutation candidates are identified in the basal trunk of the CTVT tree, but there is little evidence for ongoing positive selection. Although negative selection is detectable, its effect is largely confined to genes with known essential functions, thus implying that CTVT predominantly evolves via neutral processes.

CONCLUSION

We have traced the evolution of a transmissible cancer over several thousand years, tracking its spread across continents and contrasting the mutational processes and selective forces that moulded its genome with those described in human cancers. The identification of a highly context-specific mutational process that operated in the past but subsequently vanished, as

144 well as correlation of ultraviolet light-induced DNA damage with latitude, highlight the potential
145 for long-lived, widespread clonal organisms to act as biomarkers for mutagenic exposures.
146 Our results suggest that neutral genetic drift is the dominant evolutionary force operating on
147 cancer over the long-term, in contrast to the ongoing positive selection which is often observed
148 in short-lived human cancers. The weakness of negative selection in this asexual lineage may
149 be expected to lead to the progressive accumulation of deleterious mutations, invoking
150 Muller's Ratchet and raising the possibility that CTVT may be declining in fitness despite its
151 global success.
152

153 **Abstract**
154 The canine transmissible venereal tumour (CTVT) is a cancer lineage that arose several
155 millennia ago and survives by 'metastasising' between hosts via cell transfer. The somatic
156 mutations in this cancer record its phylogeography and evolutionary history. We constructed
157 a time-resolved phylogeny from 546 CTVT exomes and describe the lineage's worldwide
158 expansion. Examining variation in mutational exposure, we identify a highly context-specific
159 mutational process that operated early but subsequently vanished, correlate ultraviolet-light
160 mutagenesis with tumour latitude, and describe tumours with heritable hyperactivity of an
161 endogenous mutational process. CTVT displays little evidence of ongoing positive selection,
162 and negative selection is detectable only in essential genes. We illustrate how long-lived clonal
163 organisms capture changing mutagenic environments, and reveal that neutral drift is the
164 dominant feature of long-term cancer evolution.
165

166 **Introduction**
167 Transmissible cancers are malignant somatic cell clones that spread between individuals via
168 direct transfer of living cancer cells. Analogous to the metastasis of cancer to distant tissues
169 within a single body, transmissible cancers 'metastasise' as allogeneic grafts between
170 individuals within a population (*1*). Such clones have been observed only eight times in nature,
171 suggesting that they arise rarely; however, once established, transmissible cancers can
172 spread rapidly and widely and persist through time (*1, 2*). Such cancers provide a unique
173 opportunity to explore the evolution of cancer over the long-term, and to track the unusual
174 biological transition from multicellular organism to obligate conspecific asexual parasite.
175

176 The canine transmissible venereal tumour (CTVT) is the oldest and most prolific known
177 contagious cancer (*2, 3*). It is a sexually transmitted clone that manifests as genital tumours
178 in dogs. This cancer first arose from the somatic cells of an individual 'founder dog' that lived
179 several thousand years ago (*2*). The cancer survived beyond the death of this original host by
180 transfer of cancer cells to new hosts. Subsequently, this cancer has spread around the world,

181 and is a common disease in dog populations globally, although it declined and largely
182 disappeared from many Western countries during the twentieth century due to the
183 management and removal of free-roaming dogs (*4*).
184
185 Similar to cancers that remain in a single individual, CTVT accumulates somatic mutations.
186 These result from the activities of endogenous and exogenous mutational processes, and
187 genetically imprint a cancer's history of mutagenic exposures (*5*). Thus, the CTVT genome
188 can be considered a living biomarker that records the changing mutagenic environments
189 experienced by this cancer throughout millennia and across continents. Although most
190 somatic mutations in cancer have no functional effect and are considered neutral 'passenger'
191 mutations, a subset of mutations are positively selected 'driver' mutations that confer the
192 proliferation and survival advantages that spur cancer growth (*6*). Ordinary cancers, which
193 remain in a single host, often acquire additional driver mutations during tumour progression
194 (*7*); however, it is unknown whether transmissible cancers that survive for hundreds or
195 thousands of years similarly continue to adapt. It seems possible that the evolution of long-
196 lived cancers such as CTVT may instead be dominated by negative selection acting to remove
197 deleterious mutations. Finally, in addition to recording a history of exposures and signatures
198 of selection, somatic mutations provide a tool for tracing CTVT phylogeography, potentially
199 revealing how dogs, together with humans, moved around the world over the last centuries.
200 Here, we use somatic mutations extracted from the protein-coding genomes (exomes) of 546
201 globally distributed CTVT tumours to trace the history, spread, diversity, mutational exposures
202 and evolution of the CTVT clone.
203
204 **CTVT phylogeny**
205 We sequenced the exomes (43.6 megabases, Mb; mean sequencing depth ~132×) of 546
206 CTVT tumours collected between 2003 and 2016 from 43 countries across all inhabited
207 continents (Data sets S1 and S2). Candidate somatic mutations were defined as single
208 nucleotide variants (SNVs) or short insertions and deletions (indels) identified in one or more
209 CTVT tumours, but not found in 495 normal dog exomes from the CTVT tumours' matched
210 hosts. This approach yielded 160,207 variants (148,030 SNVs, 3,392 per Mb; 12,177 indels,
211 279 per Mb; Table S1). The features of this set, including its variant allele fraction distribution,
212 phylogenetic structure, comparison with the distribution of private germline variants in the dog
213 population, mutational signature composition, and non-synonymous to synonymous mutation
214 ratio (details in (*8*)), suggest that it is very highly enriched for somatic mutations. However,
215 some minimal germline variation may remain, possibly including rare germline variants from
216 the founder dog and residual contaminating alleles from matched hosts.
217

We identified the subset of the candidate somatic mutations belonging to a clock-like mutational process (specifically, cytosine-to-thymine (C>T) substitutions at CpG sites (*8, 9*)), and used these to construct a time-resolved phylogenetic tree for the CTVT lineage (Fig. 1A). The mutation rate was inferred by applying a Bayesian Poisson model to previously ascertained empirical observations (*10*), and was estimated as 6.87×10^{-7} C>T mutations per CpG site per year (*8*). The topology of the CTVT phylogenetic tree reveals a long basal trunk (Fig. 1A), representing the chain of CTVT transmissions from its origin ~6,220 years ago (95% highest posterior density interval, HPDI, 4,148–8,508 years ago) to the earliest detected node ~1,938 years ago (95% HPDI 993–3,055 years ago). This node splits a set of five tumours collected in India from the remaining population (groups labelled 57 and 58; Fig. 1A). The second and third most basal nodes (respectively ~1,004 years ago, 95% HPDI 497–1,570 years ago, and ~829 years ago, 95% HPDI 424–1,310 years ago) separate sixteen tumours from Eastern Europe and the Black Sea region, and three tumours from Northern India, from the remaining set, respectively (groups labelled 54–56 and 1; Fig. 1A). Together with evidence that the founder dog shared ancestry with ancient dog remains recovered in North-East Siberia and North America (*10*), the CTVT phylogeny supports a model whereby CTVT originated ~4,000–8,500 years ago in Central or Northern Asia, and remained within the area for the subsequent 2,000–6,000 years. Starting less than ~2,000 years ago, CTVT escaped from its founding population, perhaps due to contact between previously isolated dog groups, and spread to several locations in Asia and Europe (Fig. 1B).

The more recent history of CTVT is marked by rapid global expansion (*11*) (Figs. 1C and S1). CTVT was introduced to the Americas with early colonial contact (~500 years ago, 95% HPDI 284–888 years ago), probably initially to Central America, and further into North and South America (red sublineage 1; Fig. 1, A and C). About 300 years ago, this sublineage spread out of the Americas in an almost polytomous global sweep which brought CTVT into Africa at least five times and re-introduced the disease to Europe and Asia (black sublineage 1; Fig. 1, A and C). In parallel, a second tumour sublineage spread out of Asia or Europe into Australia and the Pacific (sublineage 2; Fig. 1, A and D). This second sublineage is also detected in North America, and its tumours were introduced to Africa on at least two occasions. By ~100 years ago, CTVT was present in dog populations worldwide, establishing local lineages that have since remained largely *in situ*. The CTVT phylogeny thus suggests that dogs, together with their neoplastic parasites, were extensively transported around the world in the fifteenth to early twentieth centuries, probably via sea travel.

Mutational processes in CTVT

254 The CTVT mutational spectrum, a representation of the six substitution types together with
255 their immediate 5' and 3' base contexts, is dominated by C>T mutations, as previously
256 described (*12, 13*) (Fig. 2A). Applying Markov chain Monte Carlo sampling on a Bayesian
257 model of mutational signatures (*8, 14*), we extracted signatures of five mutational processes
258 from the CTVT mutation load. These include three signatures that closely resemble COSMIC
259 (*15*) signatures 1, 5 and 7 (Fig. 2B). These signatures, which have previously been described
260 in CTVT (*12*), reflect endogenous mutational processes (signatures 1 and 5) and exposure to
261 ultraviolet light (UV, signature 7) (*5*). A fourth signature displaying some similarity (cosine
262 similarity 0.81) to COSMIC signature 2, which is associated with activity of APOBEC enzymes
263 (*5*), was also detected (labelled signature 2*, Fig. 2B).
264
265 The fifth signature extracted from CTVT does not resemble any previously described
266 mutational pattern. This signature, which we designate signature A, is characterised by C>T
267 mutations at N<u>C</u>C contexts and shows striking pentanucleotide sequence preference for
268 GT<u>C</u>CA (TG<u>G</u>AC on the complementary strand; Figs. 2, B and C, and S2). This extended
269 sequence preference is markedly more pronounced than previously reported pentanucleotide
270 context biases, such as those associated with UV light or DNA polymerase epsilon deficiency
271 (Fig. 2C) (*16-18*), and is not explained by the sequence composition of the canine exome (Fig.
272 S3). It is possible that signature A's causative mutagen is highly context-specific, or,
273 alternatively, that this signature's associated repair processes are ineffective at certain
274 sequence contexts ('repair shielding') (*19*). In addition, signature A displays strong
275 transcriptional strand bias, with more mutations of guanine on the untranscribed compared to
276 the transcribed strand of genes, indicating that its causative lesion is likely a guanine adduct
277 subject to transcription-coupled repair (TCR). Interestingly, the guanine-directed
278 transcriptional strand bias of signature A at T<u>C</u>C contexts counteracts the cytosine-directed
279 transcriptional strand bias of signature 7 at T<u>C</u>C, such that no overall transcriptional strand
280 bias is observed at this context in the CTVT mutational spectrum (Fig. 2A).
281
282 Using the CTVT phylogenetic tree to isolate subsets of mutations, we explored variation in
283 mutational signature exposure across time and space (Figs. S4 and S5, and Data set S3).
284 Remarkably, this revealed that signature A was highly active prior to ~2,000 years ago
285 (causing ~35% of mutations in the basal trunk of the tree, branch A1), and persisted in parallel
286 at lower levels in the two basal branches after the first node (~12% and ~9% of mutations in
287 branches A2 and A3, respectively), but then abruptly vanished (Figs. 2C and S5). Importantly,
288 signature A is not detectable within the germline of a global population of 495 dogs (Fig. S6).
289 It is possible that signature A reflects the activity of an exogenous mutagen that was uniquely
290 present in the environment that CTVT inhabited prior to its escape from its founding

population. Alternatively, it is plausible that signature A may result from an endogenous DNA-damaging agent that occurred in CTVT cells early during the lineage's history, but which ceased to accumulate from ~1,000 years ago, perhaps due to a cellular metabolic change. Although the nature of such a change is unknown, the replacement of possibly defective mitochondrial DNA by horizontal transfer, which likely occurred in parallel in branches A2 and A3 within the last ~1,690 years (*11*), may have altered the metabolic environment within CTVT cells.

Although CTVT usually occurs within the internal genital tract, it may sometimes protrude from the genital orifice or spread to perineal skin, resulting in sporadic exposure to solar UV radiation (*12, 13*). The amount of UV radiation reaching the Earth, however, varies significantly across global environments (*20*). We investigated whether latitude influenced the degree of UV exposure in CTVT tumours by estimating signature 7 contribution within subsets of mutations acquired at known latitudes. Indeed, qualitative assessment of mutational spectra of location-specific CTVT mutation subsets suggests substantial variation in UV exposure; for example, the mutational spectra of tumours collected in Mauritius show considerably more evidence of signature 7 compared with those of tumours collected in Russia (Fig. S4). Using CC>TT dinucleotide mutations (*21*) as a proxy for signature 7 (Fig. S7), we identified a non-linear association between latitude and UV exposure (Spearman's correlation –0.40, 95% HPDI [–0.65, –0.14]; Fig. 2D). By fitting CC>TT mutations observed in the basal trunk of the CTVT tree to this curve, we estimated the latitude of the CTVT founder population (Fig. 2, D and E) (*8*).

Examining the contribution of signature 5 across the CTVT lineage, we observed three independent phylogenetic groups of tumours that appear to have acquired signature 5-hyperactivity phenotypes (groups labelled 12–16, 20 and 40; Figs. 2, F and G, S4 and S5). In one case, involving tumours collected in several South and Central American countries (groups 12–16), the phenotype has been maintained for ~150 years. This phenotype is likely to result from signature 5, and not from the double-strand DNA repair deficiency-mediated COSMIC signature 3, which presents a similar mutational profile (*5, 22*), as we failed to observe the enrichment for indels which co-occurs with signature 3 (*22, 23*). It is, however, possible that these tumours were exposed to another, as yet undescribed, mutational process. Signature 5 is widespread in cancer and normal tissues and has unknown aetiology, although it may be partly associated with endogenously generated adducts subject to nucleotide excision repair (*5, 9, 18*). We annotated non-synonymous mutations occurring in the three groups' respective clonal ancestors, providing a catalogue of genes which may play a role in generation or suppression of signature 5 (Data set S4).

328

329 **CTVT mutations and gene expression**

330 The prevalence of substitution mutations in CTVT decreases with increasing gene expression,
331 likely reflecting the activity of TCR operating on DNA damage associated with signatures 7
332 and A, as well as a signature 1 preference for genes with lower expression (*16, 24, 25*) (Fig.
333 S8, A and B). We observed that exons have a higher substitution prevalence than introns,
334 possibly due to sequence context (Figs. S8A and S9). The prevalence of indels is positively
335 correlated with increasing gene expression, as has been observed in human cancers, and
336 may reflect transcription-associated damage (*26*) (Fig. S8A).

337

338 We assessed the contribution of TCR in two temporally distinct subsets of mutations: those
339 acquired prior to the earliest detectable node in the phylogenetic tree (~8,500–2,000 years
340 ago; branch A1 in Fig. 1A), and those acquired subsequent to this node (~2,000 years ago to
341 present). Interestingly, although C>T mutations acquired at T<u>C</u>C contexts in highly expressed
342 genes in branch A1 have little strand bias, likely due to the opposing transcriptional strand
343 preferences of signatures 7 and A at this context, those genes with very low expression
344 predominantly show the transcriptional strand bias associated with signature A (Fig. S8C).
345 Assuming that the transcriptional strand bias observed in these low-expressed genes reflects
346 earlier expression and subsequent silencing of genes, this suggests that there may have been
347 an early period in CTVT evolution when the lineage was exposed to signature A more intensely
348 than it was to signature 7. This may reflect variation in the climate or environment to which
349 CTVT was exposed early in its history.

350

351 **Selection in CTVT**

352 CTVT has a massive mutation burden, which exceeds that observed in even the most highly
353 mutated human cancer types (Fig. 3A). Each CTVT tumour carries on average 37,800 SNVs
354 across its predominantly diploid (*12*) exome (~2 million SNVs genome-wide; Table S2).
355 Indeed, the tally of somatic mutations that have accumulated in CTVT since it departed its
356 original host is comparable with the number of germline variants that distinguish some pairs
357 of outbred dogs (Fig. S10). Within the set of 546 tumours, 14,412 (~73%) protein-coding genes
358 carry at least one non-synonymous mutation, and 5,704 (~29%) have mutations predicted to
359 cause protein truncation (Fig. 3B).

360

361 We searched for evidence of positive selection in CTVT. The driver mutations which initially
362 caused CTVT, and which promoted its transmissible phenotype, will have occurred in the
363 basal trunk of the CTVT tree. *SETD2*, *CDKN2A*, *MYC* (previously described (*12*)), *PTEN* and
364 *RB1*, known cancer genes that frequently harbour driver mutations in human cancers (*15*),

365 carry biallelic loss-of-function or potential activating mutations in the trunk and may be early
366 drivers of CTVT (Fig. 3C and Table S3). To search for late drivers, which may have been
367 acquired in more recent parallel CTVT lineages, we identified independent mutations that
368 occurred repeatedly across the tree, and measured the normalised ratio of non-synonymous
369 to synonymous mutations (dN/dS) per gene after correcting for mutational biases and context
370 effects (*8*). This approach only yielded two uncharacterised genes with dN/dS > 1 (q-value <
371 0.05), predicted to encode a neuroligin precursor and a roundabout homologue (Data set S5).
372 The potential for these genes to act as late drivers in CTVT cannot be assessed, and it is
373 possible that local sequence structures may result in higher than expected recurrent mutation
374 rates at these loci (*27*). Overall, we find little evidence that CTVT is continuing to adapt to its
375 environment.
376
377 Negative selection, which acts to remove deleterious mutations, is very weak in human
378 cancers (*17, 28, 29*). Human cancers have short life-spans, and their evolution is dominated
379 by sweeps of strong positive selection, thus reducing the potential for negative selection to act
380 (*17*). Given its long life-span, high mutation burden and lack of ongoing positive selection, it is
381 possible that negative selection may be a more dominant force in CTVT evolution. Further,
382 unlike in ordinary cancers, in CTVT inter-tumour competition may offer more opportunities for
383 negative selection to manifest, purging lineages less able to infect new hosts and spread
384 through the host population. Indeed, negative selection has been detected operating on CTVT
385 mitochondrial genomes (*11*). Our analysis of dN/dS in CTVT across all genes, however,
386 yielded dN/dS ≈ 1 for both missense and nonsense mutations, indicating near-neutral
387 evolution (Fig. 3D and Data set S5). Similarly, dN/dS did not differ from neutrality in genes
388 categorised by expression level (Fig. 3D). Negative selection, acting both on missense and
389 nonsense mutations, could be detected, however, in sets of genes with known essential
390 functions (Fig. 3D), and was particularly pronounced for nonsense mutations in essential
391 genes occurring in haploid regions (dN/dS = 0.33, p-value < 10^{-4}). A slight signal of negative
392 selection acting on nonsense mutations in haploid regions (dN/dS = 0.88, p-value = 0.027) is
393 explained by 269 essential genes, as negative selection was not detected after removal of
394 these genes (Fig. 3D and Data set S5). These results imply that CTVT largely evolves via
395 neutral genetic drift. This may partly reflect functional obsolescence of many mammalian
396 genes in this relatively simple parasitic cancer, as well as the buffering effect of CTVT's largely
397 diploid genome (*12*). However, it is also likely that transmission bottlenecks between hosts
398 render weak selection inefficient. This may be expected to lead to the progressive
399 accumulation of deleterious mutations in the population (Muller's ratchet) (*30*), raising the
400 possibility that CTVT may be declining in fitness despite its global success.
401

402 **Discussion**

403 Studies of cancer evolution typically focus on how malignant clones alter during the first years,
404 or perhaps decades, of their existence. We have tracked the evolution of a cancer over several
405 thousand years, and compared the mutational processes and selective forces that moulded
406 its genome with those described in short-lived human cancers.

407

408 Our results suggest that neutral genetic drift may be the dominant evolutionary force operating
409 on cancer over the long-term, in contrast to the ongoing positive selection which is often
410 observed in human cancers (*7, 17*). Thus, our results suggest that CTVT may have optimised
411 its adaptation to the transmissible cancer niche early. Subsequently acquired advantageous
412 mutations may have offered incremental change of minimal benefit, such that they were
413 insufficient to overcome the neutral effects of drift. Importantly, since the 1980s, CTVT has
414 been routinely treated with vincristine, a cytotoxic microtubule inhibitor (*31*). Despite the strong
415 selection pressure imposed by vincristine treatment, we find no evidence of convergent
416 evolution of vincristine resistance mechanisms in CTVT at the level of point mutations or
417 indels.

418

419 The mechanisms whereby CTVT is tolerated by the host immune system, despite its status
420 as an allogeneic graft, are poorly understood (*32, 33*). The weakness of negative selection
421 beyond genes essential for cell viability implies that there are negligible selective pressures
422 imposed via immunoediting of somatic neoepitopes at a genome-wide level. This is perhaps
423 unsurprising, given the massive antigenic burden already presented by allogeneic epitopes.
424 These findings support evidence that CTVT largely circumvents the adaptive immune system,
425 at least during its initial stages of progressive tumour growth, perhaps in part via down-
426 regulation of major histocompatibility complex molecules (*13, 33-35*).

427

428 Our analyses reveal a mutational signature, signature A, which occurred in the past, but
429 ceased to be active from about 1,000 years ago. Interestingly, a recent study (*36*) detected
430 evidence for an excess of C>T mutations at T<u>C</u>C contexts, the mutation type most prevalent
431 in signature A, accumulating in the human germline between 15,000 and 2,000 years ago. If
432 this human mutation pulse is due to signature A, it could indicate a shared environmental
433 exposure which was once widespread, but which has now disappeared. However, we find no
434 evidence of an excess of C>T mutations at GT<u>C</u>CA pentanucleotides in the dog germline,
435 suggesting that dogs as a whole were not systemically exposed to signature A in their past.
436 Further research will be required to elucidate the biological origin of signature A and the
437 mechanism of its striking pentanucleotide sequence bias; however, this study highlights the

438 potential for long-lived, widespread clonal organisms to act as biomarkers for the activity of
439 mutational processes.
440
441 Genomic instability and ongoing positive selection are often considered key hallmarks of
442 carcinogenesis (*37*). CTVT does not have an intrinsically high point mutation rate ('genomic
443 instability'), at least at the level of SNVs, and its vast mutation burden simply reflects the
444 lineage's age. We find no clear evidence for continued positive selection beyond initial truncal
445 events. Thus, CTVT illustrates that, once spawned and sufficiently well-adapted to its niche,
446 neither hallmark is necessary to sustain cancer over the long term.
447
448 CTVT is a remarkable biological entity. It is the oldest, most prolific and most divergent cancer
449 lineage known in nature; it has spread throughout the globe and has seeded its tumours in
450 many thousands of dogs. Here, we have traced this cancer's route through the steppes of
451 Asia and Europe and as an unwelcome stowaway on global voyages. We have observed the
452 patterns in its mutational profiles reflecting the dynamics of its exogenous and endogenous
453 environment. Further, we have shown that CTVT largely evolves via neutral processes, and
454 that the mutations that it continues to acquire may pose a threat, rather than an advantage, to
455 its long-term fitness.
456
457 **REFERENCES AND NOTES**

458 1. M. J. Metzger, S. P. Goff, A Sixth Modality of Infectious Disease: Contagious Cancer
459 from Devils to Clams and Beyond. *PLoS Pathog* **12**, e1005904 (2016).
460 2. A. Strakova, E. P. Murchison, The cancer which survived: insights from the genome
461 of an 11000 year-old cancer. *Curr Opin Genet Dev* **30**, 49-55 (2015).
462 3. D. P. Blaine, *A domestic treatise on the diseases of horses and dogs*. (T. Boosey,
463 London, 1810).
464 4. A. Strakova, E. P. Murchison, The changing global distribution and prevalence of
465 canine transmissible venereal tumour. *BMC Vet Res* **10**, 168 (2014).
466 5. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer. *Nature*
467 **500**, 415-421 (2013).
468 6. M. R. Stratton, P. J. Campbell, P. A. Futreal, The cancer genome. *Nature* **458**, 719-
469 724 (2009).
470 7. N. McGranahan, C. Swanton, Clonal Heterogeneity and Tumor Evolution: Past,
471 Present, and the Future. *Cell* **168**, 613-628 (2017).
472 8. Materials and methods are available as supplementary materials.
473 9. L. B. Alexandrov *et al.*, Clock-like mutational processes in human somatic cells. *Nat*
474 *Genet* **47**, 1402-1407 (2015).

475 10. M. Ní Leathlobhair *et al.*, The evolutionary history of dogs in the Americas. *Science*
476 **361**, 81-85 (2018).
477 11. A. Strakova *et al.*, Mitochondrial genetic diversity, selection and recombination in a
478 canine transmissible cancer. *Elife* **5**, (2016).
479 12. E. P. Murchison *et al.*, Transmissible dog cancer genome reveals the origin and
480 history of an ancient cell lineage. *Science* **343**, 437-440 (2014).
481 13. B. Decker *et al.*, Comparison against 186 canid whole-genome sequences reveals
482 survival strategies of an ancient clonally transmissible canine tumor. *Genome Res*
483 **25**, 1646-1655 (2015).
484 14. K. C. Gori, A. Baez-Ortega, sigfit: flexible Bayesian inference of mutational
485 signatures. *bioRxiv* **372896**, (2018).
486 15. S. A. Forbes *et al.*, COSMIC: exploring the world's knowledge of somatic mutations
487 in human cancer. *Nucleic Acids Res* **43**, D805-811 (2015).
488 16. E. D. Pleasance *et al.*, A comprehensive catalogue of somatic mutations from a
489 human cancer genome. *Nature* **463**, 191-196 (2010).
490 17. I. Martincorena *et al.*, Universal Patterns of Selection in Cancer and Somatic
491 Tissues. *Cell* **171**, 1029-1041 e1021 (2017).
492 18. L. B. Alexandrov *et al.*, The Repertoire of Mutational Signatures in Human Cancer.
493 *bioRxiv*, (2018).
494 19. S. S. Lange, K. M. Vasquez, HMGB1: the jack-of-all-trades protein is a master DNA
495 repair mechanic. *Mol Carcinog* **48**, 571-580 (2009).
496 20. R. L. McKenzie, P. J. Aucamp, A. F. Bais, L. O. Bjorn, M. Ilyas, Changes in
497 biologically-active ultraviolet radiation reaching the Earth's surface. *Photochem*
498 *Photobiol Sci* **6**, 218-231 (2007).
499 21. G. P. Pfeifer, Y. H. You, A. Besaratinia, Mutations induced by ultraviolet light. *Mutat*
500 *Res* **571**, 19-31 (2005).
501 22. S. Nik-Zainal *et al.*, Landscape of somatic mutations in 560 breast cancer whole-
502 genome sequences. *Nature* **534**, 47-54 (2016).
503 23. S. Morganella *et al.*, The topography of mutational processes in breast cancer
504 genomes. *Nat Commun* **7**, 11383 (2016).
505 24. N. J. Haradhvala *et al.*, Mutational Strand Asymmetries in Cancer Genomes Reveal
506 Mechanisms of DNA Damage and Repair. *Cell* **164**, 538-549 (2016).
507 25. E. Letouze *et al.*, Mutational signatures reveal the dynamic interplay of risk factors
508 and cellular processes during liver tumorigenesis. *Nat Commun* **8**, 1315 (2017).
509 26. S. Jinks-Robertson, A. S. Bhagwat, Transcription-associated mutagenesis. *Annu Rev*
510 *Genet* **48**, 341-359 (2014).

511 27. X. Zou *et al.*, Short inverted repeats contribute to localized mutability in human
512 somatic cells. *Nucleic Acids Res* **45**, 11213-11221 (2017).
513 28. D. Weghorn, S. Sunyaev, Bayesian inference of negative and positive selection in
514 human cancers. *Nat Genet* **49**, 1785-1788 (2017).
515 29. J. Van den Eynden, S. Basu, E. Larsson, Somatic Mutation Patterns in Hemizygous
516 Genomic Regions Unveil Purifying Selection during Tumor Evolution. *PLoS Genet*
517 **12**, e1006506 (2016).
518 30. H. J. Muller, The Relation of Recombination to Mutational Advance. *Mutat Res* **106**,
519 2-9 (1964).
520 31. E. I. Amber, R. A. Henderson, J. B. Adeyanju, E. O. Gyang, Single-drug
521 chemotherapy of canine transmissible venereal tumor with cyclophosphamide,
522 methotrexate, or vincristine. *J Vet Intern Med* **4**, 144-147 (1990).
523 32. D. Frampton *et al.*, Molecular Signatures of Regression of the Canine Transmissible
524 Venereal Tumor. *Cancer Cell* **33**, 620-633 e626 (2018).
525 33. H. V. Siddle, J. Kaufman, Immunology of naturally transmissible tumours.
526 *Immunology* **144**, 11-20 (2015).
527 34. C. Murgia, J. K. Pritchard, S. Y. Kim, A. Fassati, R. A. Weiss, Clonal origin and
528 evolution of a transmissible cancer. *Cell* **126**, 477-487 (2006).
529 35. A. Fassati, N. A. Mitchison, Testing the theory of immune selection in cancers that
530 break the rules of transplantation. *Cancer Immunol Immunother* **59**, 643-651 (2010).
531 36. K. Harris, J. K. Pritchard, Rapid evolution of the human mutation spectrum. *Elife* **6**,
532 (2017).
533 37. P. C. Nowell, The clonal evolution of tumor cell populations. *Science* **194**, 23-28
534 (1976).
535 38. M. P. Hoeppner *et al.*, An improved canine genome and a comprehensive catalogue
536 of coding genes and non-coding transcripts. *PLoS One* **9**, e91172 (2014).
537 39. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler
538 transform. *Bioinformatics* **25**, 1754-1760 (2009).
539 40. A. Rimmer *et al.*, Integrating mapping-, assembly- and haplotype-based approaches
540 for calling variants in clinical sequencing applications. *Nat Genet* **46**, 912-918 (2014).
541 41. M. E. Skinner, A. V. Uzilov, L. D. Stein, C. J. Mungall, I. H. Holmes, JBrowse: a next-
542 generation genome browser. *Genome Res* **19**, 1630-1638 (2009).
543 42. D. Jones *et al.*, cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to
544 Detect Somatic Single Nucleotide Variants in NGS Data. *Current protocols in*
545 *bioinformatics* **56**, 15.10. 11 (2016).
546 43. R. Rahbari *et al.*, Timing, rates and spectra of human germline mutation. *Nat Genet*
547 **48**, 126-133 (2016).

548 44. W. McLaren *et al.*, The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122
549 (2016).
550 45. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis
551 of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
552 46. P. Skoglund, E. Ersmark, E. Palkopoulou, L. Dalen, Ancient wolf genome reveals an
553 early divergence of domestic dog ancestors and admixture into high-latitude breeds.
554 *Curr Biol* **25**, 1515-1519 (2015).
555 47. L. R. Botigue *et al.*, Ancient European dog genomes reveal continuity since the Early
556 Neolithic. *Nat Commun* **8**, 16082 (2017).
557 48. B. Carpenter *et al.*, Stan: A probabilistic programming language. *Journal of statistical*
558 *software* **76**, (2017).
559 49. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with
560 BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973 (2012).
561 50. A. J. Drummond, A. Rambaut, B. Shapiro, O. G. Pybus, Bayesian coalescent
562 inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**,
563 1185-1192 (2005).
564 51. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, P. J. Campbell, M. R. Stratton,
565 Deciphering signatures of mutational processes operative in human cancer. *Cell Rep*
566 **3**, 246-259 (2013).
567 52. A. Baez-Ortega, K. Gori, Computational approaches for discovery of mutational
568 signatures in cancer. *Brief Bioinform*, (2017).
569 53. K. Suzuki, D. Tonien, K. Kurosawa, K. Toyota, Birthday paradox for multi-collisions.
570 *IEICE Transactions on Fundamentals of Electronics, Communications and Computer*
571 *Sciences* **91**, 39-45 (2008).
572 54. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and
573 bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
574 55. G. Tischler, S. Leonard, biobambam: tools for read pair collation based algorithms on
575 BAM files. *Source Code for Biology and Medicine* **9**, 13 (2014).
576 56. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq:
577 transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
578 57. J. Rainer, L. Gatto, C. X. Weichenberger, ensembldb: an R package to create and
579 use Ensembl-based annotation resources. *Bioinformatics*, (2019).
580 58. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion
581 for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
582 59. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the
583 integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat*
584 *Protoc* **4**, 1184-1191 (2009).

585 60. R. Nielsen, Z. Yang, Likelihood models for detecting positively selected amino acid
586 sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929-936 (1998).
587 61. L. D. Hurst, The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends*
588 *Genet* **18**, 486 (2002).
589 62. C. Roadmap Epigenomics *et al.*, Integrative analysis of 111 reference human
590 epigenomes. *Nature* **518**, 317-330 (2015).
591 63. T. Hart, K. R. Brown, F. Sircoulomb, R. Rottapel, J. Moffat, Measuring error rates in
592 genomic perturbation screens: gold standards for human functional genomics. *Mol*
593 *Syst Biol* **10**, 733 (2014).
594 64. V. A. Blomen *et al.*, Gene essentiality and synthetic lethality in haploid human cells.
595 *Science* **350**, 1092-1096 (2015).
596 65. T. Wang *et al.*, Identification and characterization of essential genes in the human
597 genome. *Science* **350**, 1096-1101 (2015).
598 66. J. D. Arroyo *et al.*, A Genome-wide CRISPR Death Screen Identifies Genes
599 Essential for Oxidative Phosphorylation. *Cell Metab* **24**, 875-885 (2016).
600 67. P. A. Futreal *et al.*, A census of human cancer genes. *Nat Rev Cancer* **4**, 177-183
601 (2004).
602 68. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**,
603 646-674 (2011).
604

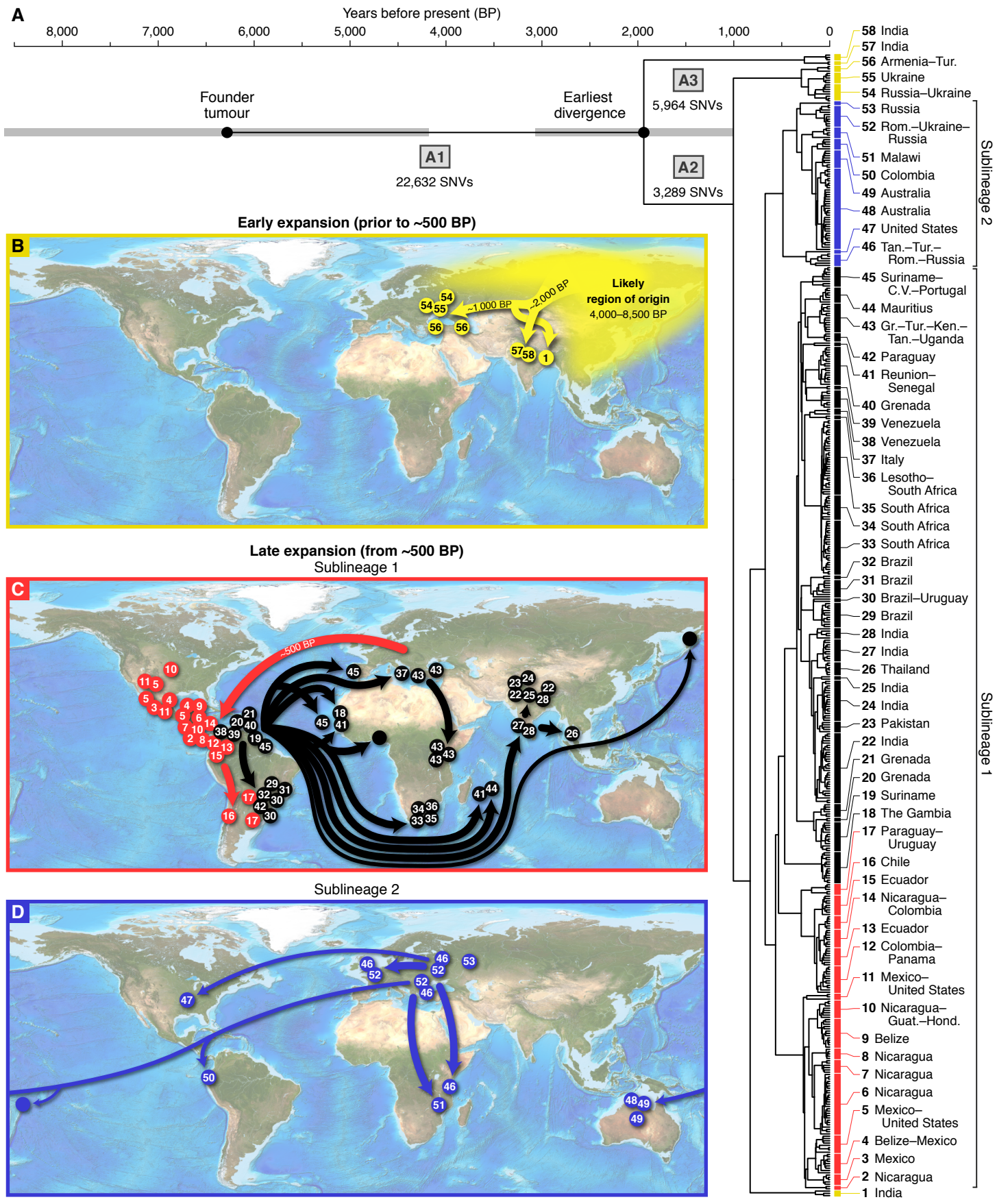
605 **ACKNOWLEDGMENTS**

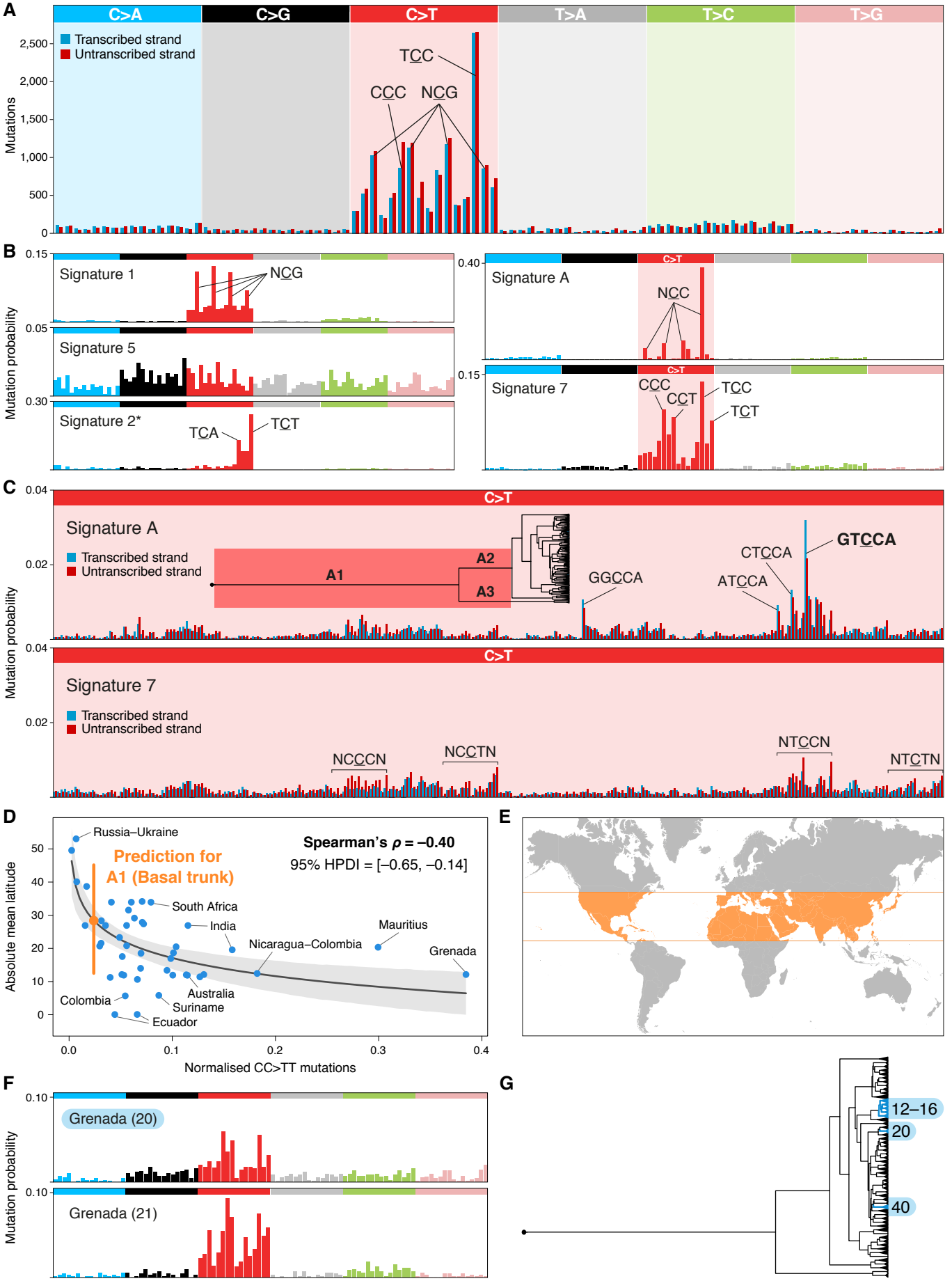
606 We acknowledge the Core Sequencing Facility, IT groups and members of the Cancer
607 Genome Project at the Wellcome Sanger Institute. We thank the following individuals for
608 useful information and for their help obtaining samples for this project: Ilona Airikkala-Otter,
609 Juliana Alzate-Ocampo, Diana Argüello, Jose Ignacio Arias, Clara Lee Arnold, Sue Barrass,
610 Ekaterina Batrakova, Rafaela Bortolotti Viéra, Nikki Brown, Fernando Constantino Casas,
611 John Cooper, Amici Cannis Cotacachi, Stephen M. Cutter, Johan de Vos, Lytvynenko Dmytro,
612 Phillip Farnham, Ariberto Fassati, Andres Fernandez-Riomalo, Ricardo Gaitan, David
613 Hanzlíček, Rafael Ricardo Huppes, John M. Igundu, Matilde Jimenez-Coello, Debra
614 Kamstock, Patrick Kelly, Tatiana Korytina, Anna Kuznetsova, Gleidice Eunice Lavalle
615 (Universidade Federal de Minas Gerais), Olakunle AbdulRasaq Lawal, Thabo Lerotholi, Marco
616 Lima-Maigua, Jimmy Loayza-Feijoo, Mayra López-Bucheli, Mwangi Maina, Margarita
617 Mancero-Albuja, Cynthia Marchiori Bueno, Luis Martínez-López, Alfredo Martínez-Meza,
618 Bedan M. Masuruli, Talita Mariana Morata Raposo, Jude Mulholland, Claudio Murgia, Alvira
619 Murison Swartz, Fran Nargi, Marsden M. Onsare, Edwin Ortiz-Rodríguez, Elisabeth Peach,
620 Lisa Pellegrini, Gerry Polton, Freddy Proaño-Pérez, Juan Carlos Ramirez-Ante, Cameron
621 Raw, Ceseltina Semedo, Ivan Stoikov, Irina Swarisch, Mirela Tinucci Costa, Emily Turitto, M.

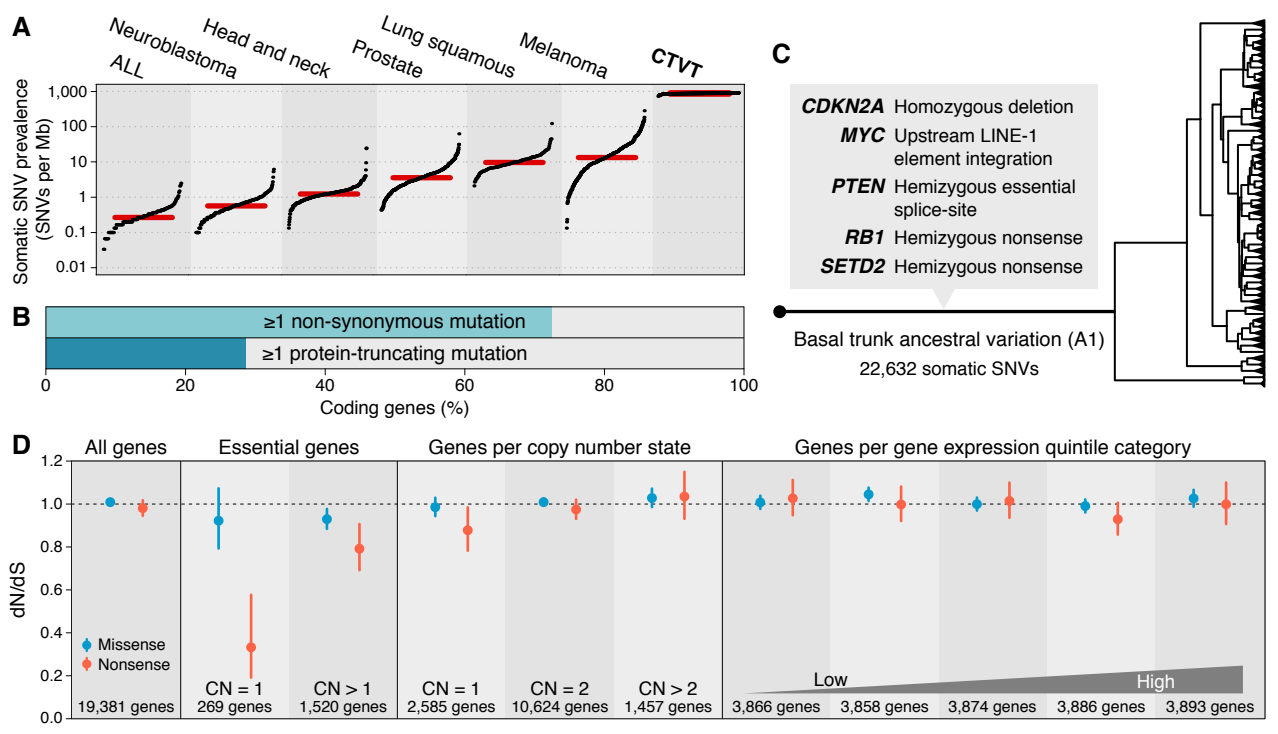
Rifat Vural, David Walker, Robin Weiss, Kevin Xie, Maurice Zandvliet, staff at Animal Medical Centre Belize City (Belize), veterinary surgeons and staff at Help in Suffering (Jaipur, India), staff at Hopkins Belize Humane Society (Belize), veterinary workers at Pet Centre (UVAS, Lahore, Pakistan), students from St. George's University (True Blue, Grenada, West Indies) who assisted with sample collection, staff at Veterinary Clinic “El Roble” (Chile), staff and volunteers at World Vets (Gig Harbor, USA) and staff at the WVS International Training Centre in Ooty (India). We are grateful to the following organisations for helpful information: American College of Veterinary Internal Medicine (ACVIM), Animal Balance, Animal Care Association (The Gambia), Animal Management in Rural and Remote Indigenous Communities (AMRRIC), Associacao Bons Amigos de Cabo Verde, Humane Society of Cozumel, Humane Society Veterinary Medical Association–Rural Area Veterinary Services (HSVMA–RAVS), Israel Veterinary Medical Association, Italian Veterinary Oncology Society, Rural Vets South Africa, Veterinary Cancer Society, Veterinary Society of Surgical Oncology (VSSO), VetPharma, Vets Beyond Borders, ViDAS and Coco’s Animal Welfare, The Spanky Project, VWB/VSF Canada, West Arnhem Land Dog Health Program (WALDHeP), World Small Animal Veterinary Association (WSAVA), МИР ВЕТЕРИНАРИИ (World Veterinary Medicine). **Funding:** This work was supported by Wellcome (102942/Z/13/A) and by a Philip Leverhulme Prize from the Leverhulme Trust. A.S. was supported by a Postgraduate Student Award from the Kennel Club Charitable Trust. **Author contributions:** E.P.M. designed and directed the project. A.B.-O. developed methods and led computational data analysis. K.G. developed methods and assisted with computational analysis. A.St. collected samples, performed laboratory work, designed exome probes, oversaw sequencing and provided conceptual advice. J.L.A., K.M.A., L.B.-I., T.N.B., J.L.B., C.B., A.C.D., A.M.C., H.R.C., J.T.C., E.D., K.F.d.C., A.B.D.N., A.P.d.V., L.D.K., E.M.D., A.R.E.H., I.A.F., M.F., E.F., S.N.F, F.G.-A., O.G., P.G.G., R.F.H.M., J.J.G.P.H., R.S.H., N.I., Y.K., C.K., D.K., A.K., S.J.K., M.L.-P., M.L., A.M.L.Q., T.L., G.M., S.M.C., M.F.M.-L., M.M., E.J.M., B.N., K.N., W.N., S.J.N., A.O.-P., F.P.-O., M.C.P., K.P., R.J.P., J.F.R., J.R.G., H.S., S.K.S., O.S., A.G.S., A.E.S.-S., A.Sv., L.J.T.M., I.T.N., C.G.T., E.M.T., M.G.v.d.W., B.A.V., S.A.V., O.W., A.S.W.-M. and S.A.E.W. provided clinical samples. Y.-M.K., M.N.L. and M.S. assisted with analysis and contributed to interpretation of results. J.W. contributed to sample management and curation. M.R.S., L.B.A. and I.M. provided technical advice and assisted with interpretation of results. A.B.-O. and E.P.M. wrote the manuscript and designed the figures. All authors commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Whole-exome sequence data have been deposited in the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) under overarching accession number ERP109580. Variant calling data and other data supporting analyses have been deposited in the University of Cambridge Repository (https://www.repository.cam.ac.uk) under DOI

659 10.17863/CAM.24962. Custom algorithms employed for data processing and analysis are
660 available in GitHub (https://github.com/baezortega/TCG2019).
661
662 **SUPPLEMENTARY MATERIALS**
663 Materials and Methods
664 Figs. S1 to S16
665 Tables S1 to S4
666 Data sets S1 to S6
667 References (*38–68*)
668
669 **Main figure legends**
670 **Fig. 1. Phylogeny and geographical expansion of CTVT.** (**A**) Time-resolved phylogenetic
671 tree inferred from clock-like exonic somatic variation in CTVT. Each tip is a tumour and
672 sampling locations are labelled. Numbers refer to phylogenetic groups displayed on maps in
673 **B**–**D**. Sublineages 1 and 2, referred to in **C** and **D** respectively, are marked. Three groups of
674 ancestral somatic variation (A1, A2, A3) and their respective numbers of single nucleotide
675 variants (SNVs) are indicated. The estimated age of the CTVT founder tumour and the earliest
676 detected node are indicated in years before present (BP), with grey error bars depicting
677 Bayesian 95% HPDI. (**B** to **D**) Maps presenting likely routes of early (prior to ~500 years BP)
678 and late (from ~500 years BP) expansion of CTVT. Numbered circles indicate the
679 geographical locations of phylogenetic groups labelled in **A**; arrows represent inferred
680 geographical movements. Circle and arrow colours indicate different sets of geographical
681 movements, as labelled in **A**. Thin arrows indicate expansion routes for which there is limited
682 phylogenetic evidence; dots without numbers denote tumours that are not represented in the
683 tree. C.V., Cape Verde; Gr., Greece; Guat., Guatemala; Hond., Honduras; Ken., Kenya; Rom.,
684 Romania; Tan., Tanzania; Tur., Turkey.
685
686 **Fig. 2. Mutational processes in CTVT.** (**A**) Trinucleotide-context mutational spectrum of
687 somatic SNVs in a single CTVT tumour. Horizontal axis presents 96 mutation types displayed
688 in pyrimidine context. Relevant trinucleotide mutation contexts are indicated. (**B**) Trinucleotide-
689 context mutational spectra of extracted mutational signatures 1, 5, 2*, A and 7, with relevant
690 trinucleotide mutation contexts indicated. (**C**) Pentanucleotide-context mutational spectra of
691 signature A (top) and signature 7 (bottom). Horizontal axis presents 256 C>T mutation types
692 with relevant mutation contexts indicated. The inset tree shows the phylogenetic branches
693 with exposure to signature A. (**D**) Bayesian logarithmic regression and Spearman's correlation
694 between absolute mean latitude and normalised CC>TT mutations in phylogenetic groups
695 shown in Fig. 1A. Normalised CC>TT mutations represent the ratio between group-unique

696 CC>TT mutations and group-unique C>T changes at CpG dinucleotides. The black line and
697 shadowed area indicate the regression curve and associated 95% HPDI. The orange dot and
698 bars represent predicted absolute mean latitude and associated 90% prediction interval for
699 the basal trunk ancestral variation (group A1). Posterior median and 95% HPDI of the
700 correlation coefficient are shown. (**E**) Map showing the latitude range corresponding to the
701 90% prediction interval for group A1, presented in **D**, in the northern hemisphere. (**F**)
702 Trinucleotide-context mutational spectra of a phylogenetic tumour group showing evidence of
703 signature 5 hyperactivity (top) and a closely related group without signature 5 hyperactivity
704 (bottom). (**G**) Diagram indicating the phylogenetic situation of the tumour groups displaying
705 signature 5 hyperactivity.
706
707 **Fig. 3. Selection in CTVT.** (**A**) Somatic SNV prevalence across six human cancer types and CTVT.
708 Dots represent individual tumours; red lines indicate median SNV prevalence. ALL, acute
709 lymphoblastic leukaemia. (**B**) Bars showing the percentage of protein-coding genes in the CTVT
710 genome harbouring ≥1 non-synonymous somatic mutation (SNV or indel; 14,412 genes) and ≥1
711 somatic protein-truncating somatic mutation (5,704 genes). (**C**) Diagram presenting the putative
712 driver events found in the set of basal trunk ancestral variants (group A1, Fig. 1A). A description of
713 each somatic alteration is shown next to the corresponding gene symbol. (**D**) Exome-wide dN/dS
714 ratios estimated for somatic SNVs in all protein-coding genes (left) and in sets of genes defined
715 according to gene essentiality, copy number state and expression level. Estimates of dN/dS are
716 presented for missense (blue) and nonsense (orange) mutations in each gene group. The dashed
717 line indicates dN/dS = 1 (neutrality); error bars indicate 95% confidence intervals.


A
Years before present (BP)
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
Founder tumour
Earliest divergence
A1
22,632 SNVs
A2
3,289 SNVs
A3
5,964 SNVs
Early expansion (prior to ~500 BP)
B
~1,000 BP
~2,000 BP
Likely region of origin
4,000–8,500 BP
Late expansion (from ~500 BP)
Sublineage 1
C
~500 BP
Sublineage 2
D
58 India
57 India
56 Armenia–Tur.
55 Ukraine
54 Russia–Ukraine
53 Russia
52 Rom.–Ukraine–Russia
51 Malawi
50 Colombia
49 Australia
48 Australia
47 United States
46 Tan.–Tur.–Rom.–Russia
45 Suriname–C.V.–Portugal
44 Mauritius
43 Gr.–Tur.–Ken.–Tan.–Uganda
42 Paraguay
41 Reunion–Senegal
40 Grenada
39 Venezuela
38 Venezuela
37 Italy
36 Lesotho–South Africa
35 South Africa
34 South Africa
33 South Africa
32 Brazil
31 Brazil
30 Brazil–Uruguay
29 Brazil
28 India
27 India
26 Thailand
25 India
24 India
23 Pakistan
22 India
21 Grenada
20 Grenada
19 Suriname
18 The Gambia
17 Paraguay–Uruguay
16 Chile
15 Ecuador
14 Nicaragua–Colombia
13 Ecuador
12 Colombia–Panama
11 Mexico–United States
10 Nicaragua–Guat.–Hond.
9 Belize
8 Nicaragua
7 Nicaragua
6 Nicaragua
5 Mexico–United States
4 Belize–Mexico
3 Mexico
2 Nicaragua
1 India
Sublineage 2
Sublineage 1


A
C>A
C>G
C>T
T>A
T>C
T>G
Transcribed strand
Untranscribed strand
Mutations
TCC
CCC
NCG
B
Signature 1
NCG
Signature 5
Signature 2*
TCA
TCT
Mutation probability
Signature A
NCC
Signature 7
CCC
CCT
TCC
TCT
C
Signature A
A1
A2
A3
GGCCA
CTCCA
ATCCA
GTCCA
Signature 7
NCCCN
NCCTN
NTCCN
NTCTN
D
Russia–Ukraine
Prediction for
A1 (Basal trunk)
Spearman's ρ = –0.40
95% HPDI = [–0.65, –0.14]
South Africa
India
Mauritius
Nicaragua–Colombia
Grenada
Australia
Suriname
Colombia
Ecuador
Absolute mean latitude
Normalised CC>TT mutations
E
F
Grenada (20)
Grenada (21)
G
12–16
20
40


A
ALL
Neuroblastoma
Head and neck
Prostate
Lung squamous
Melanoma
CTVT
Somatic SNV prevalence (SNVs per Mb)
1,000
100
10
1
0.1
0.01
B
≥1 non-synonymous mutation
≥1 protein-truncating mutation
0
20
40
60
80
100
Coding genes (%)
C
CDKN2A Homozygous deletion
MYC Upstream LINE-1 element integration
PTEN Hemizygous essential splice-site
RB1 Hemizygous nonsense
SETD2 Hemizygous nonsense
Basal trunk ancestral variation (A1)
22,632 somatic SNVs
D
All genes
Essential genes
Genes per copy number state
Genes per gene expression quintile category
dN/dS
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Missense
Nonsense
19,381 genes
CN = 1
269 genes
CN > 1
1,520 genes
CN = 1
2,585 genes
CN = 2
10,624 genes
CN > 2
1,457 genes
Low
High
3,866 genes
3,858 genes
3,874 genes
3,886 genes
3,893 genes