

1 **Genetically determining individualized clinical reference ranges for the biomarker tryptase**
2 **can limit unnecessary procedures and unmask myeloid neoplasms**

3
4 **Authors:** Jack Chovanec¹, Ilker Tunc², Jason Hughes³, Joseph Halstead⁴, Allyson Mateja⁵, Yihui
5 Liu¹, Michael P. O’Connell¹, Jiwon Kim¹, Young Hwan Park¹, Qinlu Wang⁶, Quang Le⁷, Mehdi
6 Pirooznia², Neil N. Trivedi⁸, Yun Bai⁹, Yuzhi Yin⁹, Amy P. Hsu¹⁰, Josh McElwee¹¹, Sheryce
7 Lassiter¹², Celeste Nelson¹, Judy Bando¹, Thomas DiMaggio¹³, Julij Šelb¹⁴, Matija Rijavec¹⁴,
8 Melody C. Carter⁹, Hirsh D. Komarow⁹, Vito Sabato¹⁵, Joshua Steinberg¹⁶, Kurt M. Hafer¹⁷,
9 Elizabeth Feuille¹⁸, Christopher S. Hourigan¹⁹, Justin Lack²⁰, Paneez Khoury²¹, Irina Maric²²,
10 Roberta Zanotti²³, Patrizia Bonadonna²⁴, Lawrence B. Schwartz⁷, Joshua D. Milner²⁵, Sarah C.
11 Glover²⁶, Didier G. Ebo¹⁵, Peter Korošec¹⁴, George H. Caughey⁸, Erica H. Brittain²⁷, Ben Busby²⁸,
12 Dean D. Metcalfe⁹, and Jonathan J. Lyons.^{1†}

13
14 **Affiliations:** ¹Translational Allergic Immunopathology Unit, Laboratory of Allergic Diseases,
15 National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD;
16 ²Bioinformatics and Computational Biology Core, National Heart, Lung, and Blood Institute,
17 National Institutes of Health, Bethesda, MD; ³Foundation Medicine, Cambridge, MA;
18 ⁴Birmingham Women's and Children's NHS Foundation Trust, Birmingham, England; ⁵Clinical
19 Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research,
20 Frederick, MD; ⁶Bioinformatics and Computational Biosciences Branch, Office of Cyber
21 Infrastructure and Computational Biology, NIAID, NIH, Bethesda, MD; ⁷Department of Internal
22 Medicine, Virginia Commonwealth University, Richmond, Virginia; ⁸Cardiovascular Research
23 Institute and Department of Medicine, University of California San Francisco, San Francisco, CA,

24 and Veterans Affairs Medical Center, San Francisco, CA; ⁹Mast Cell Biology Section, Laboratory
25 of Allergic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of
26 Health, Bethesda, MD; ¹⁰Immunopathogenesis Section, Laboratory of Clinical Immunology and
27 Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health,
28 Bethesda, MD; ¹¹Nimbus Therapeutics, Cambridge, MA; ¹²Clinical Research Directorate,
29 Frederick National Laboratory for Cancer Research, Frederick, MD; ¹³Fungal Pathogenesis
30 Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and
31 Infectious Diseases, National Institutes of Health, Bethesda, MD; ¹⁴University Clinic of
32 Respiratory and Allergic Diseases, Golnik, Slovenia; ¹⁵Department of Immunology, Allergology,
33 and Rheumatology, University of Antwerp and Antwerp University Hospital, Infla-Med, Centre
34 of excellence, University of Antwerp, Antwerp, Belgium; ¹⁶Division of Allergy and Clinical
35 Immunology, Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI;
36 ¹⁷Department of Medicine, Stanford University, Palo Alto, CA; ¹⁸Division of Allergy and Clinical
37 Immunology, Department of Pediatrics, Weill Cornell Medical College, New York, NY;
38 ¹⁹Myeloid Malignancies Section, National Heart, Lung, and Blood Institute, National Institutes of
39 Health, Bethesda, MD; ²⁰NIAID Collaborative Bioinformatics Resource (NCBR), National
40 Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD; ²¹Human
41 Eosinophil Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious
42 Diseases, National Institutes of Health, Bethesda, MD; ²²Hematology Service, Department of
43 Laboratory Medicine, Clinical Center, National Institutes of Health, Bethesda, MD; ²³Department
44 of Medicine, Section of Hematology, Verona University Hospital, Verona, Italy; ²⁴Allergy Unit,
45 Verona University Hospital, Verona, Italy; ²⁵Division of Allergy, Immunology and Rheumatology,
46 Columbia University, New York, NY; ²⁶Division of Digestive Diseases, Department of Medicine,

47 University of Mississippi Medical Center, Jackson, MS; ²⁷Biostatistics Research Branch, National
48 Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD;
49 ²⁸National Center for Biotechnology Information, National Library of Medicine, National
50 Institutes of Health, Bethesda, MD

51

52 †**Correspondence:** Jonathan J. Lyons, M.D.

53 9000 Rockville Pike,

54 Building 29B, Room 5NN18, MSC 1889

55 Bethesda, MD 20892

56 Phone: (301) 443-5250

57 Fax: (301) 480-5757

58 Email: jonathan.lyons@nih.gov

59

60 **Funding:** This project has been funded in whole or in part with federal funds from the Division of
61 Intramural Research of the National Institute of Allergy and Infectious Diseases, NIH and with
62 federal funds from the National Cancer Institute, National Institutes of Health, under Contract Nos.
63 HHSN261201500003I and 75N910D00024. This project was also funded in part with federal
64 funds from the Division of Intramural Research of the National Heart, Lung, and Blood Institute,
65 NIH. The content of this publication does not necessarily reflect the views or policies of the
66 Department of Health and Human Services, nor does mention of trade names, commercial
67 products, or organizations imply endorsement by the U.S. Government. The involvement of BB
68 and JH was supported by the Intramural Research Program of the National Library of Medicine.
69 SCG is supported by US Public Health Service Grant NIH R21DE028378. VS is a Senior Clinical

70 Researcher of the Research Foundation Flanders/ Fonds Wetenschappelijk Onderzoek (FWO:
71 1804518N). DE is a Senior Clinical Researcher of the Research Foundation Flanders/Fonds
72 Wetenschappelijk Onderzoek (FWO: 1800614N).

73

74 **Conflict of interest statement:** VCU receives royalties from Thermo Fisher for their tryptase test
75 that are shared with LBS as its inventor. None of the remaining authors have relevant conflicts of
76 interest to report.

77 **ABSTRACT**

78 Serum tryptase is a biomarker used to aid in the identification of certain myeloid neoplasms, most
79 notably systemic mastocytosis, where baseline (BST) levels >20 ng/mL are a minor criterion for
80 diagnosis. Whereas clonal myeloid neoplasms are rare, the common cause for elevated BST is the
81 genetic trait hereditary alpha-tryptasemia (H α T) caused by increased germline *TPSAB1* copy
82 number. To date, the precise structural variation and mechanism(s) underlying elevated BST in
83 H α T and the general clinical utility of tryptase genotyping, remain undefined. Through cloning,
84 long-read sequencing, and assembling of the human tryptase locus from an individual with H α T,
85 and validating our findings *in vitro* and *in silico*, we demonstrate that BST elevations arise from
86 over-expression of replicated *TPSAB1* loci encoding wild-type α -tryptase due to co-inheritance of
87 a linked over-active promoter element. Modeling BST levels based upon *TPSAB1* replication
88 number we generate new individualized clinical reference values for the upper limit of 'normal'.
89 Using this personalized laboratory medicine approach, we demonstrate the clinical utility of
90 tryptase genotyping, finding that in the absence of H α T, BST levels >11.4 ng/mL frequently
91 identify indolent clonal mast cell disease. Moreover, substantial BST elevations (e.g., >100 ng/mL)
92 which would ordinarily prompt bone marrow biopsy, can result from *TPSAB1* replications alone
93 and thus be within 'normal' limits for certain individuals with H α T.

94 INTRODUCTION

95 The tryptase locus is structurally complex in humans (Table 1). Approximately two-thirds of
96 people have α -tryptase, encoded at *TPSABI* on one or both alleles, whereas everyone has β -
97 tryptase encoded at one or both *TPSB2* alleles, as well as at non- α -tryptase encoding *TPSABI* loci
98 (1-3). Copy number gain and loss of tryptase gene sequences encoding both α - and β -tryptases
99 have also been reported (4). However, the structures of such copy number variants (CNV) remain
100 unknown. The most common CNVs observed among Western populations are *TPSABI* gene
101 replications encoding α -tryptase, a genetic trait known as hereditary alpha-tryptasemia (H α T) (5).
102 H α T is inherited in an autosomal dominant pattern and has been shown to affect nearly 6% of the
103 general population in the U.S., U.K., and E.U (6-9). *De novo* replications have not been reported.
104 *TPSABI* replications are associated with elevated basal serum tryptase (BST) levels of at least 2-
105 3-fold higher than median BST levels of healthy individuals without H α T (10). While it has been
106 hypothesized that additional *TPSABI* copies identified in H α T are present within the tryptase locus
107 at chromosome 16p13.3, and that over-expression of α -tryptase yields the disproportionate
108 increases in BST levels seen relative to copy number (5), this has never been demonstrated.
109 Moreover, since only a small region of *TPSABI* has been probed to demonstrate gene replication
110 (5, 11), it remains unknown whether extra-allelic copies of *TPSABI* encode wild-type or novel
111 sequences containing homology with α -tryptase at the probe site.

112 Whereas H α T is the most common heritable cause for elevated BST, the clonal mast cell
113 disorder systemic mastocytosis (SM) is a common acquired cause (10, 12), and BST levels are
114 used routinely as a biomarker to screen for this disorder. Currently, the WHO criteria define a BST
115 level >20 ng/mL as a minor criterion for the clinical diagnosis of SM (13) and this cut-off is
116 frequently used in clinical decision making as an indication for bone marrow biopsy in

117 symptomatic individuals (14). However, approximately one out of every four individuals with
118 H α T has a BST level >20 ng/mL, representing an estimated 7.5 million people in the U.S. alone
119 (4, 15). Because H α T has been shown to augment immediate hypersensitivity symptoms in a
120 number of conditions (7, 8), many of these individuals are likely to undergo unwarranted invasive
121 work-up for SM, including bone marrow biopsy, if tryptase genotype is not taken into account.
122 Conversely, because H α T appears to account for most elevations in BST, observed elevations in
123 patients who do not have H α T could be considerably enriched for other pathologies - most notably
124 myeloid neoplasms - warranting workup when such elevations in BST may be modest (i.e., <20
125 ng/mL).

126 To address these questions, we cloned, long-read sequenced, and assembled the human
127 tryptase locus containing a *TPSABI* replication, finding the coding sequence to be wild-type α -
128 tryptase. However, a series of unique proximal non-coding variants were also identified that
129 distinguished replicated (α^{DUP} -tryptase) from non-replicated (α^{WT} -tryptase) sequences. An
130 expanded DNA motif within the 5'-UTR was also linked to *TPSABI* replication-associated variants
131 and demonstrated increased *in vitro* promoter activity relative to the paralogous region in the non-
132 replicated promoter. Using *in vitro* and *in silico* analyses of RNA sequences, we confirmed the
133 relative over-expression of α^{DUP} -tryptase sequences in primary basophils, cultured mast cells, and
134 publicly available RNA-sequence datasets. Applying this knowledge, we generated a new genetic-
135 based model for BST clinical reference ranges based upon *TPSABI* replication number and
136 examined the potential real-world impact of coupling tryptase genotyping with these newly
137 defined values in the work-up of patients with clonal mast cell disorders. The potentially practice-
138 changing implications for identification of indolent clonal mast cell disease, the diagnosis of SM,
139 and the elimination of unnecessary bone marrow biopsies are discussed herein.

140 **RESULTS**

141 *Increased TPSAB1 copy number occurs at the tryptase locus and encodes wild-type alpha-*
142 *tryptase*

143 A bacterial artificial chromosome (BAC) library was generated from an individual homozygous
144 for *TPSAB1* duplications (genotype $\alpha\alpha/\beta:\alpha\alpha/\beta$) and a clone containing the complete human
145 tryptase locus was identified. Using single molecule real time (SMRT) technology, we sequenced
146 and subsequently assembled the locus *de novo* (Fig. 1a), allowing study of the structure and
147 sequence of this variant-containing allele for the first time. As anticipated, the assembly displayed
148 marked structural dissimilarity with the reference human genome (GRCh37/hg19) but did not
149 contain novel sequence (Fig. S1, and online data repository). The additional *TPSAB1* copy
150 encoding α -tryptase (α^{DUP}) was located within the tryptase locus mapping to 16p13.3 between
151 *TPSB2* and the non-replicated *TPSAB1* locus, as had been suggested based upon prior *in vitro*
152 experiments (5). The coding sequence was identical to that of the adjacent wild-type α -tryptase
153 sequence (α^{WT}) (Table S4; online repository), and both sequences were in reverse orientation.

154 The sequence at *TPSB2* in our BAC clone was identified as β 1-tryptase. While β 1-tryptase
155 at *TPSB2* is not common, it has been reported in approximately 5% of individuals screened from
156 the HapMap cohort (16). Given that this prevalence was comparable to the prevalence of H α T
157 reported in Western populations, we obtained the samples reported to contain β 1-tryptase at *TPSB2*
158 in order to determine whether they were all individuals with H α T. However, only 4 out of 15 of
159 these individuals (27%) with β 1-tryptase at *TPSB2* were found to have H α T. Thus, β 1-tryptase at
160 *TPSB2* is not exclusive to H α T.

161

162 *TPSAB1* replications encoding α^{DUP} -tryptase are linked to an expanded promoter

163 Elevated BST levels among patients with H α T have been shown to result from increased
164 constitutive release of pro-tryptases (5, 17). To determine whether this increase results from over-
165 expression of α^{DUP} -tryptase transcripts, we compared non-coding sequences at the two *TPSAB1*
166 loci and identified an expanded (~800 bp) repetitive DNA motif approximately 1 kilobase (kB) 5'
167 of the consensus transcription start site for α^{DUP} -tryptase (Fig. 1b). This sequence was
168 approximately 700 bp longer than the paralogous region of the α^{WT} promoter containing 27
169 repeated and 3 repeated sequences, respectively (Fig. 1c). Repeat regions at the same 5' location
170 were also observed in publicly available assemblies of α^{WT} sequences and confirmed by gel
171 electrophoresis in amplified α^{WT} -tryptase and α^{DUP} -tryptase promoters from 9 unrelated
172 individuals with H α T (Fig. S2). Subsequent cloning and Sanger sequencing of three additional
173 α^{WT} and α^{DUP} -tryptase promoters confirmed these relative differences (Fig. S3). Together these
174 data demonstrate an expanded DNA repeat motif in the promoter of *TPSAB1* replications encoding
175 α^{DUP} -tryptase that is consistently co-inherited.

176

177 ***Increased basal α^{DUP} -tryptase expression is associated with elevated basal promoter activity***

178 To quantify the functional outcome of the expanded promoter region linked to α^{DUP} -tryptase
179 sequences we next interrogated total α - and β -tryptase as well as α^{WT} - and α^{DUP} -tryptase transcript
180 expression arising from *TPSAB1* and *TPSB2*. Cultured primary human mast cells and isolated
181 basophils from individuals with H α T expressed higher levels of total α -tryptase transcripts, and
182 over-expressed α^{DUP} -tryptase sequences relative to other tryptase isoforms (Fig. 1d,e).
183 Importantly, total α -tryptase gene expression levels were several orders of magnitude higher than
184 those of β -tryptases, even when they were present in allelic balance (e.g., 2 α ,2 β versus 3 α ,3 β).

185 To determine if the expanded 5' sequences uniquely associated with α^{DUP} -tryptase
186 contributed to the relative over-expression of α^{DUP} - to α^{WT} -tryptase, we next cloned the promoters
187 containing these paralogous regions into fluorescent reporter plasmids using our BAC assembly
188 and other SMRT-based *de novo* assemblies of the locus as sequence references (see methods for
189 referenced assemblies). These constructs were sequence-verified and transfected into MonoMac-
190 6 cells – an AML line that expresses tryptases (18, 19). The expanded α^{DUP} -tryptase promoter
191 demonstrated significantly greater basal activity relative to the α^{WT} -tryptase promoter (Fig. 2a).
192 This *in vitro* finding correlated directly with BST levels among individuals with H α T where the
193 absolute number of *TPSAB1* replications was found to best correlate with BST level regardless of
194 tryptase genotype derived from *TPSAB1* and *TPSB2* (Fig. 2b). Therefore, elevated BST in H α T
195 results from over-expression of α^{DUP} -tryptase at replicated *TPSAB1* loci linked to a promoter with
196 increased basal activity.

197

198 ***A unique haplotype demonstrates conservation of the identified over-active promoter at***
199 ***replicated TPSAB1 loci***

200 To examine the generalizability of our finding of an expanded promoter linked to α^{DUP} -tryptase
201 at *TPSAB1*, we again examined proximal non-coding sequences associated with different tryptase
202 isoforms and identified five unique 5' variants – two substitutions in intron 1, two in the 5' UTR,
203 and one in the proximal promoter (Fig. 1b). Using these variants, we defined haplotypes, that when
204 combined with coding sequences, uniquely distinguished α^{DUP} -tryptase (CACCT) from α^{WT} -
205 tryptase (GGTTC), β 1-tryptase (GACCC), and δ -tryptase (GGTTC) in our BAC assembly (Fig.
206 1b; see Table S1 for complete contextual sequences).

207 Next, we examined publicly available genome assemblies to determine if these haplotypes
208 were conserved. Critically, *TPSAB1* and *TPSB2* have been shown to be in nearly complete linkage
209 disequilibrium and are co-inherited in virtually all individuals such that major haplotypes exist and
210 inheritance can be inferred (16, 20). The major tryptase isoforms encoded at *TPSAB1* are β 1-
211 tryptase followed by α^{WT} -tryptase, and at *TPSB2* are β 3-tryptase and β 2-tryptase, when on β/β and
212 α/β alleles respectively. As expected we observed GACCC linked to β 1-tryptase at *TPSB2* on
213 β 1/ β 3 (β/β) alleles, and GGTTTC linked to α^{WT} -tryptase sequences at *TPSAB1* on α^{WT}/β 2 (α/β)
214 alleles in all 13 queried assemblies as well as well as in the GRCh37/hg19 reference assembly
215 (Table 1). Additional isoform-specific haplotypes at these two loci - GACCT linked to β 2-tryptase
216 at the *TPSB2* locus only when α^{WT}/β 2 (α/β) alleles were present and CACCT linked to β 3-
217 tryptases at the *TPSB2* locus only when β 1/ β 3 (β/β) alleles were present - were also identified.
218 Neither the α^{DUP} -tryptase-specific promoter expansion nor the linked CACCT haplotype were
219 observed in association with α^{WT} -tryptase sequences in any of the seven assemblies containing
220 α^{WT}/β 2 (α/β) (Table 1).

221 Using these haplotypes, we next re-examined genome sequence reads from 183 individuals
222 without increased *TPSAB1* copy number, and from 32 affected individuals from 10 families with
223 H α T (5). Employing a paired-read approach, we found that GACCC (β 1-tryptase) and CACCT
224 (β 3-tryptase) were present in association with 98% (324/331) of individuals with β/β alleles.
225 Moreover, the CACCT haplotype consistently segregated with the β/β haplotype in the 16 families
226 without H α T in whom tryptase sequence inheritance could be examined. Whereas all α/β
227 haplotypes (n = 6) were associated with the α^{WT} -tryptase-linked GGTTTC and β 2-tryptase-linked
228 GACCT (Table 2), the CACCT haplotype linked to α^{DUP} -tryptase in our BAC clone, segregated

229 universally with *TPSAB1* replications in all 32 individuals with H α T. The GACCC haplotype –
230 linked to β 1-tryptase most often present at *TPSAB1* but present at *TPSB2* in our BAC clone – was
231 also found to be universally present among individuals with H α T and co-segregated with α^{DUP} -
232 tryptase containing alleles suggesting that β 1-tryptase at *TPSB2* is frequently present on alleles
233 containing *TPSAB1* replications, as seen in our BAC clone (Table 2). Therefore, these *in silico*
234 findings suggest conservation of the unique promoter haplotypes we identified in our BAC
235 assembly, and that over-expression of α^{DUP} -tryptase could be a generalizable phenomenon.

236 To confirm these findings at scale *in vitro*, we had to overcome the repetitive and GC-rich
237 nature of the region. To do so we developed a ddPCR assay capable of detecting the most proximal
238 variant which distinguished the α^{DUP} -tryptase promoter haplotype (CACCTI) from that of α^{WT} -
239 tryptase (GGTTC) (see Fig. S4 for description and representative data). The corresponding variant
240 listed in the reference genome (GRCh37/hg19 16:1290818C>T) has only been reported in
241 association with *TPSB2* – ostensibly linked to dominant isoform β 3-tryptase at this locus.
242 However, the reverse primer in our assay was designed to only hybridize with α -tryptase
243 sequences.

244 The CACCTI variant was confirmed to be universally present in linkage with α^{DUP} -tryptase
245 in 120 affected individuals with increased *TPSAB1* copy number from 101 families. Conversely,
246 CACCTI was never found in linkage with α^{WT} -tryptase in 81 unaffected individuals from 69
247 families who did not have increased *TPSAB1* copy number but carried one or two α^{WT} -tryptase-
248 encoding copies (Fig. 2c). Furthermore, the ratios of identified promoter copy number confirmed
249 that all 166 H α T-associated α^{DUP} -tryptase sequences on 156 alleles were linked to the C>T variant
250 (Fig. 2d) and by inference, the expanded promoter.

251 Interestingly, in 30/120 individuals with H α T, α^{WT} -tryptase was not linked to GGTTC; 10
252 of these α^{WT} -tryptase promoters appeared to contain the C>T missense, and the remaining 19
253 alleles may have had this or another unknown missense variant or indel, the latter of which may
254 have resulted in failed probe or primer hybridization. Importantly, our *in silico* analysis identified
255 a GGTTT haplotype linked to α^{WT} -tryptase in one family with H α T. In a second family a single
256 base-pair deletion in the -6 position from C>T was observed, which we have labelled GGTT(-
257 6delC)C was also identified (Table 2). This deletion interfered with probe hybridization in our
258 assay potentially accounting for some of the 19 α^{WT} -tryptase containing alleles with ambiguous
259 promoter linkage. The α^{WT} -tryptase promoter from these samples also failed to amplify and was
260 not successfully cloned, suggesting yet another possible α^{WT} -tryptase promoter may exist. Indeed,
261 public database interrogation yielded at least six different α -tryptase sequences exhibiting a high
262 degree of homology (Table S4). Regardless, α^{WT} -tryptase containing alleles with undefined
263 promoter haplotypes, or even those linked to the C>T intronic variant, were not associated BST
264 differences *in vivo* (Fig. S5), indicating that these undefined promoters are not associated with
265 differences in gene expression at linked α^{WT} -tryptase encoding *TPSABI* loci. Collectively, these
266 data support conservation of a unique haplotype linked to an over-active promoter element at
267 replicated *TPSABI* loci that is associated with over-expression of α -tryptase by mast cells and
268 basophils.

269

270 ***α^{DUP} -Tryptase transcripts are over-expressed in public datasets***

271 Using α - and β -tryptase isoform-specific coding variants and the uniquely linked proximal non-
272 coding haplotypes, a 39-base consensus sequence was defined that contained four variants which
273 could distinguish α^{WT} -, α^{DUP} -, and β -tryptase transcripts; given the proximity of the first variant

274 to the transcriptional start site, only the distal three variants were used for read alignment (Fig. 1b,
275 Table S2). Querying 58 non-disease-associated transcriptome datasets, 863/4160 samples were
276 identified with exact transcript matches to ≥ 1 of the three defined 39-bp consensus sequences.
277 α^{DUP} -Tryptase was observed in 3% (25/863) and α^{WT} -tryptase transcripts were identified in 34%
278 of individuals (289/863) (Table S5). Because linked meta-data were limited, the race and
279 ethnicities of these samples are unknown. While Caucasians are likely to be over-represented in
280 these samples, other ethnicities and racial groups are likely included which may explain the lower-
281 than-expected prevalence of H α T (3% v. 5.6%).

282 Using the same method, we next examined the expression levels of α^{WT} -, α^{DUP} -, and β -
283 tryptase transcripts in these and other public datasets. This analysis confirmed over-expression of
284 both α^{WT} -, α^{DUP} -tryptase transcripts relative to β -tryptase (Fig. 3a). While over-expression of α^{WT} -
285 tryptase was also observed in our primary gene expression data, it is possible that some δ -tryptase
286 reads could align to the α^{WT} -tryptase consensus sequence, due to haplotype sequence homology,
287 but not to α^{DUP} -tryptase where the linked haplotype is not observed in linkage with δ -tryptase.
288 Therefore, misalignment of δ -tryptase reads could lead to a potential over-estimation of α^{WT} -
289 tryptase expression levels. To exclude this possibility, we examined BST levels – which only
290 measure total secreted α - and β -tryptases – among individuals without H α T. As reported (21), we
291 saw a modest but significant positive correlation between BST levels and increasing α^{WT} -tryptase
292 copy number at *TPSAB1* (Fig. 3b). Thus, consistent with the RNA sequence read alignment results,
293 α^{WT} does appear to be modestly over-expressed relative to β -tryptases. Because of the variable
294 levels of total tryptase transcript in any given sample, we next employed regression analysis to
295 quantify the relative expression of α^{DUP} -tryptase to α^{WT} -tryptase more precisely. Consistent with
296 all our prior data, α^{DUP} -tryptase was found to be over-expressed relative to α^{WT} -tryptase by a factor

297 of 5.2-fold (95% CI 4.864 to 5.499) in these datasets ($R^2=0.8$, $P<0.0001$) (Fig. 3c). Thus, *in silico*,
298 *in vitro*, and *in vivo* findings collectively indicate that elevated BST in H α T results from basal
299 over-expression of wild-type α -tryptase at replicated *TPSABI* loci linked to co-inherited over-
300 active promoter elements.

301
302 ***Modeling BST levels based upon TPSABI replication number redefines clinical reference***
303 ***ranges***

304 Using this knowledge, we created a data-driven model from 204 individuals with normal *TPSABI*
305 copy number and 309 individuals with H α T in whom there was no clinical evidence of clonal mast
306 cell disease (MCD), to generate prediction intervals with upper limits for predicted BST levels
307 based upon *TPSABI* replication number (Fig. 3d) and created an online application for clinical
308 use: **B**asal **S**erum **T**ryptase **C**linical cut-off **A**ssigned by **L**ocus **C**opy number of **U**TR-**L**inked
309 element and **A**ssociated **T***TPSABI* **E**ncoded **R**eplication (BST CALCULATOR) available at:
310 <https://bst-calculator.niaid.nih.gov/>. These thresholds can be used to determine where a bone
311 marrow or further clinical evaluation may be needed for individuals with or without H α T when
312 another clinical indication for such evaluation may be lacking.

313 This model not only redefined clinically meaningful upper limits for BST levels among
314 individuals with H α T, but it also for the first time defines a clinically actionable upper limit among
315 individuals without H α T as 11.4 ng/mL (Table 3). Remarkably, this has been the upper limit of
316 normal for BST commonly used in most clinical laboratories base. Moreover, based upon the data,
317 this cut-off of 11.4 ng/mL would appear to be valid in defining a clinically abnormal elevated BST
318 when considering the diagnosis of clonal mast cell disease - as opposed to the currently accepted
319 20 ng/mL - provided that H α T has been ruled out by tryptase genotyping.

320 This model fit the primary data well ($R^2=0.76$, $P<0.001$) with only one individual being
321 identified in the entire primary dataset as outside the 99.5% threshold (Fig. 3d; Table 3; Fig. S6)
322 and consistent with in vitro and in vivo data, demonstrated linearity on the log-transformed scale
323 (Fig. S6). Interestingly, this individual who did not conform to the model, had provocative clinical
324 phenotype which included chronic spontaneous urticaria and angioedema as well as syncopal
325 episodes with alcohol consumption that had prompted a bone marrow biopsy prior to genotyping.
326 While bone marrow sections demonstrated left-shifted megaloblastoid erythroid hyperplasia, focal
327 areas with increased pronormoblasts, and occasional dyserythropoietic forms, allele-specific PCR
328 for *KIT* p.D816V was negative, no mast cell aggregates or aberrant markers (i.e., CD2, CD25)
329 were identified, and no definitive diagnosis was made.

330 Finally to prospectively validate the modeled cut-off clinically, we measured BST levels
331 in an additional 130 individuals with H α T caused by *TPSAB1* duplications in whom BST levels
332 were unknown and no other clinical indication for bone marrow biopsy was present. In all 130
333 prospectively examined individuals, BST levels fell below the 99.5% cut-off (median 15.1 ng/mL,
334 range 8.0-35.0 ng/mL).

335

336 ***Application of tryptase genotyping and modeled BST levels improves biomarker utility***

337 To further examine the real-world implications of applying this new model clinically, we analyzed
338 all samples sent or referred to the NIH clinical center for tryptase genotyping from 2016-2021;
339 377 of these samples had not been used to construct the model. None of the participants had kidney
340 disease or other clinical presentations consistent with genetic disorders known to be associated
341 with elevated BST (13). Of these, 409 samples were identified with BST >11.4 ng/mL; 285 of
342 these had H α T, 98 had MCD, 21 had both H α T and MCD, 2 individuals had hypereosinophilic

343 disorders - one of whom also had H α T - and 4 individuals did not have a clear diagnosis at the
344 time of genotyping. Thus, 98.8% (404/409) of individuals with BST >11.4 ng/mL could be
345 accounted for by H α T and/or MCD (Fig. 3e). Moreover, 97.3% (255/262) of samples with BST
346 levels between 8-11.4 ng/mL were associated with H α T and/or MCD, suggesting that even BST
347 >8 ng/mL is uncommon among individuals without these conditions. Consistent with these data is
348 the modeled 95% threshold among individuals without H α T or MCD of 7.99 ng/mL, indicating
349 that less than 5% of the total population would be predicted to have BST >8 ng/mL.

350 The 5 individuals with BST >11.4 ng/mL who did not have H α T or evidence of MCD
351 underwent paired peripheral blood exome and deep exome sequencing (300x) of granulocyte and
352 mononuclear bone marrow aspirate fractions. In all five individuals, variants in genes commonly
353 mutated in hematological malignancies (e.g., *TET2*, *RUNX1*, *CEBPA*, *MAP2K1*, *NOTCH2*,
354 *KMT2C*) were identified (Table S6); in one individual this was associated with idiopathic
355 hypereosinophilic syndrome. Together with those individuals meeting criteria for MCD, these data
356 demonstrate that BST >11.4 ng/mL in the absence of H α T frequently identifies indolent clonal
357 mast cell disease - 95% (98/103) of our referral cohort - or more rarely identifies patients with
358 genetic variants suggestive of occult or evolving myeloid neoplasms.

359 Of note, 38.2% (55/144) samples with BST >20ng/mL were a result of H α T alone. Of
360 those 55 participants, 54 were less than their respective threshold based on replication number. If
361 tryptase genotyping was not considered, and instead the current minor criterion for SM of BST
362 >20 ng/mL was used as a sole indication for bone marrow evaluation, 54 of these individuals could
363 have been subjected to unnecessary invasive bone marrow examination. This is particularly
364 relevant among individuals with two or more *TPSAB1* replications, where median BST levels were

365 >20 ng/mL and can be >100 ng/mL, as seen in two related individuals with ten *TPSABI*
366 replications in cis (Fig. 2b, 3d; Table 3).

367 Finally, of the 21 individuals identified with both H α T and MCD (Table 4), 10 individuals
368 had indolent SM (ISM), 8 of whom had BST levels above the 99.5% upper prediction limit dictated
369 by *TPSABI* replication number. Ten individuals had BST levels below the 99.5% upper prediction
370 limit based upon *TPSABI* replication number in our model; 6 with monoclonal mast cell activation
371 syndrome (MMAS), 2 with ISM, 1 with diffuse cutaneous mastocytosis, and 1 with well-
372 differentiated ISM. All these individuals had clinical implications for a bone marrow beyond the
373 BST level and would have received a bone marrow regardless. Only one of these individuals - with
374 the rare phenotype of well-differentiated ISM - would no longer have met clinical criteria for their
375 diagnosis if the upper prediction limit was used as a minor criterion for the diagnosis of SM (rather
376 than the current level of >20 ng/mL). Together, these data demonstrate that where tryptase
377 genotyping is clinically available, these new reference ranges based upon *TPSABI* replication
378 number can be applied to the work-up of patients with elevated BST, establishing robust thresholds
379 for which patients should undergo more extensive work-up including bone marrow aspiration and
380 biopsy regardless of clinical presentation or symptomatology, and as an individualized minor
381 criterion for the clinical diagnosis of SM.

382

383 **DISCUSSION**

384 Based upon our findings, H α T is a genetic trait best described as a naturally occurring over-
385 expression model, whereby a conserved over-active protomer is co-inherited with α -tryptase-
386 encoding *TPSABI* gene replications driving basal over-expression of α^{DUP} -tryptase that when
387 translated, is indistinguishable from other α -tryptase proteins (i.e., α^{WT} -tryptase). Using this

388 knowledge and clinical data, we have generated a model that redefines clinically meaningful cut-
389 offs for the upper limit of normal for serum tryptase levels based upon tryptase genotype.
390 Moreover, we have demonstrated the clinical utility of using these individualized values in the
391 evaluation of patients with elevated serum tryptase levels and in the diagnosis of systemic
392 mastocytosis.

393 We acknowledge that MCD is not the only clonal myeloid neoplasm associated with
394 elevated BST, rather it is the predominant one referred to our centers. Moreover, several studies
395 have now demonstrated an increased prevalence of H α T among symptomatic individuals with SM
396 (6-8) - likely due to potentiation of mast cell-mediated symptoms and reactions. Additional
397 prospective studies will be helpful in validating the use of these new reference values in additional
398 patients with H α T, MCD and other myeloid neoplasms. Severe renal dysfunction can also affect
399 BST levels, as can other parasitic diseases which are rare in most Western countries (15). These
400 are also not represented or evaluated in this study and should be considered when applying this
401 model to patients. However, the model was generated based upon individuals without clinical signs
402 or symptoms of these disorders. Thus, we expect these new reference values to remain robust in
403 their predictive value, as has been suggested by a recent retrospective study of a regional health
404 system where 93% (54/58) of individuals with BST >11.4 ng/mL had H α T, a myeloid neoplasm,
405 or chronic kidney disease (22).

406 The data used to generate our model is largely from individuals with two or fewer
407 replications. Thus, while available data for higher order copy number are consistent with our
408 model, and indeed were incorporated into it, the generated cut-offs at high *TPSABI* replication
409 number (e.g., ≥ 3) are largely extrapolations. The a priori evidence demonstrated a linear
410 relationship between BST and replication number providing support for this extrapolation.

411 However, there is less certainty about the upper limits of BST level levels past two *TPSABI*
412 replications, and cut-offs for individuals with higher order copy number may require refinement
413 as additional data become available.

414 To our knowledge, the application of genotypic information to determine clinical reference
415 ranges in a personalized manner has not previously been used in laboratory medicine. However,
416 many common clinical laboratory measures (e.g., immunoglobulin levels) are not normally
417 distributed, and in some cases outliers may be similarly determined by heritable traits. Thus, we
418 anticipate this kind of precision approach to clinical laboratory medicine will expand in the future
419 as next generation sequencing becomes increasingly utilized in standard clinical practice.

420

421 **METHODS**

422 **Study participants and samples**

423 Patients, family members, and healthy volunteers provided informed consent on IRB-approved
424 research protocols led by investigators with expertise in Allergy/Immunology at institutions that
425 specialize in clonal and non-clonal mast cell disorders designed to study mastocytosis, or genetic
426 diseases affecting the immune system at the NIH Clinical Center (NCT00852943, NCT01164241,
427 NCT00044122, NCT00001756, NCT007197190), Antwerp University Hospital, Belgium
428 (B300201525454), Verona University Hospital, Italy (protocol No. 39620), the University of
429 Florida (IRB 201702274), the University of Mississippi Medical Center (IRB 2019-0082), or
430 University Clinic Golnik, Slovenia (KME 150/09/13). All study participants had BST levels
431 measured and tryptase genotyping performed (N=1,178). Among participants referred for
432 evaluation or diagnosed with a mast cell-associated disorder (N=575), complete history and
433 physical examinations were performed.

434

435 **Bone marrow biopsy and aspirate**

436 Individuals presenting with signs or symptoms suggestive of a clonal mast cell disease (13) had
437 additional clinical work-up that included bone marrow biopsy and aspirate. Immunohistochemistry
438 of bone marrow sections was performed for enumeration and characterization of mast cells (KIT
439 and tryptase), and evaluation of CD2 and CD25 expression in aspirate and/or tissue section. Allele-
440 specific PCR for *KIT* p.D816V in peripheral blood and bone marrow was also performed in these
441 patients. In the rare case that an individual declined bone marrow biopsy, peripheral blood *KIT*
442 p.D816V was performed by allele-specific or ddPCR. If the result was positive, these individuals
443 were considered to have clonal mast cell disease and included. Whereas, if *KIT* p.D816V screening
444 of peripheral blood was negative in an individual with an incomplete work-up for a suspected
445 myeloid neoplasm, such individuals were excluded from the study.

446

447 **Total basal serum tryptase quantification**

448 Total basal serum tryptase (BST) levels were measured using the commercially available
449 ImmunCAP assay (Thermo Fisher Scientific, Phadia AB, Uppsala, Sweden) or ELISA as
450 described (16), performed in CLIA-certified laboratories (Mayo Clinic, Rochester, NY and
451 Virginia Commonwealth University, Richmond, VA).

452

453 **Bacterial artificial chromosome (BAC) library generation**

454 Genomic DNA was isolated from PBMC of an individual homozygous for $\alpha\alpha\beta$ tryptase alleles.
455 HMW DNA fragments were then partially digested with the restriction enzyme BamHI and size
456 selected prior to ligation of fragments into the pCC1BAC vector (Epicentre Biotechnologies,

457 Madison, WI) and transformation of DH10B *E. coli* cells (17, 18). The clones were robotically
458 chosen and replicated. Replicated clone copies were used as a source plate for constructing nylon
459 filters. The filter hybridization was carried out using the DIG method for filter hybridization as
460 described (19) (Amplicon Express, Pullman, WA). Verification of putative clones was carried out
461 by conventional and digital droplet PCR (ddPCR).

462

463 **BAC sequencing and assembly**

464 Genomic DNA from selected clones was isolated (20) and sequenced using Single-molecule real-
465 time sequencing (SMRT) sequencing technology (PacBio, Menlo Park, CA) at DNA Link (South
466 Korea). De novo assembly was accomplished using HGAP2 and the Canu assembler for PacBio
467 on SMRT Analysis platform version 2.2.0 and the SMRTPortal: SMRTAnalysis build 133377,
468 Daemon version v2.2.0 build 132105, SMRTpipe version v2.2.0 build 132739, SMRT Portal
469 version v2.2.0 build 133335, SMRT View version v2.2.0 build 132578.

470

471 **Bioinformatic analyses**

472 *Paired-read tryptase haplotype analysis*

473 Genome sequence reads mapping to the tryptase locus were re-aligned to a 142-bp region (Table
474 S1). Detailed alignments for each read were parsed to extract the sequence at each of the five
475 specified variable positions if covered by the read. Haplotype assignments were determined based
476 on the collection of positional sequence assignments determined for each read pair. Generated
477 haplotypes were then correlated to tryptase genotype and linkage to inherited alleles was confirmed
478 when pedigrees were available.

479

480 ***RNA-seq dataset analyses***

481 Using a 39-bp consensus sequence (Table S2), gene expression datasets (see Accession codes for
482 list of datasets) were reanalyzed using NCBI Magic-BLAST (<https://ncbi.github.io/magicblast/>),
483 in order to map reads to 3 gene isoforms: $\alpha^{\text{DUP-}}$, $\alpha^{\text{WT-}}$, and β -tryptase. Samples were designated
484 as tryptase-positive, tryptase-negative, $\alpha^{\text{DUP-}}$, $\alpha^{\text{WT-}}$, and β -tryptase only based on the criteria of
485 having qualified reads that mapped exactly without any mismatches to the interval between 25 bp-
486 45 bp from 5' end of the transcripts (see Supplemental Methods).

487

488 **Promoter amplification and cloning**

489 Promoter amplicons were generated from genomic DNA from individuals with increased *TPSABI*
490 copy number using primers designed on conserved sequences present in all identified alpha- and
491 beta-tryptase sequences (FWD – GGGCAAGTCCACAGGGAGCT; REV –
492 CTGGGGAGCAAGGAGGAGCA) in order to amplify all sequences between the ATG start site
493 and a conserved region ~1-2 kb 5' of the variably expanded repeat region (Fig. 1). Amplification
494 was confirmed by gel electrophoresis and clones of products were generated using the TOPO[®]
495 Cloning Kit (Thermo, Waltham, MA) and transformed into One Shot[®] TOP10 Chemically
496 Competent *E. coli* (Thermo). Single colonies were selected, confirmed to contain single intact
497 clones by PCR and gel electrophoresis, and Sanger sequenced using the FWD and REV, as well
498 as two additional internal primers (TGCAGGTGCAACCCCAGGA and
499 TCCTGGGGTTGCACCTGCA).

500 For $\alpha^{\text{DUP-}}$ and $\alpha^{\text{WT-}}$ -tryptase-specific promoter cloning, the reverse primer
501 (GACGATACCCGCTTGCTGCAG) was located across alpha-tryptase-specific exonic sequence
502 used for the isoform-specific genotyping assay, and for $\alpha^{\text{WT-}}$ -tryptase the universal FWD primer

503 was used and for α^{DUP} -tryptase a different sequence normally only seen linked to beta-tryptase
504 was used (GGGCAAGTCCACAGGGAGCT). Resulting amplicon size was confirmed by gel
505 electrophoresis (Fig. S2).

506

507 **Reporter assay**

508 Sequence-validated clones corresponding to α^{DUP} - and α^{WT} -tryptase promoters identified in the
509 BAC assembly were subcloned into the reporter plasmid pDD-AmCyan1 (Takara Bio USA, Inc.,
510 Mountain View, CA) and Sanger sequence verified. MonoMac-6 cells were transfected with the
511 reporter plasmid pDD-AmCyan1 containing α^{DUP} - or α^{WT} -tryptase promoter clones by
512 electroporation using Cell Line Nucleofector™ Kit V (Lonza, Basel, Switzerland) using the setting
513 U-005 according to manufacturer instructions. Cells were cultured in standard media in the
514 presence of Shield1 (Takara Bio) according to manufacturer instructions and basal fluorescence
515 was measured at indicated time points and recorded using an LSR Fortessa (BD Biosciences) and
516 analyzed using FlowJo (Treestar, Ashland, OR).

517

518 **Basophil isolation**

519 Following isolation of PBMC via density centrifugation, basophils were negatively selected using
520 the MACS® Diamond Basophil Isolation Kit (Miltenyi Biotec, San Diego, CA).

521

522 **Cell culture**

523 *Primary mast cells*

524 Peripheral blood mononuclear cells (PBMCs) were isolated using density centrifugation and
525 CD34⁺ cells were positively selected using the ferromagnetic bead-based MACS® system

526 (Miltenyi) and the number of cells was quantified. Mast cells were then differentiated under the
527 established conditions described (21).

528

529 ***Cell lines***

530 MonoMac-6 cells were cultured in RPMI 1640 + 10% FBS + 2 mM L-glutamine + non-essential
531 amino acids + 1 mM sodium pyruvate + 10 µg/ml human insulin).

532

533 **Droplet digital PCR (ddPCR)**

534 ***Tryptase genotyping***

535 Alpha- and beta-tryptase sequences at *TPSAB1* and *TPSB2* were genotyped as described (5) using
536 droplet digital PCR (ddPCR). Briefly, genomic DNA was extracted from PBMCs, cell lines, or
537 obtained from the HapMap biorepository and restriction endonuclease treated. Custom
538 primer/probe sets specific for alpha- and beta-tryptase were then employed using the PrimePCR
539 ddPCR copy number reference *AP3B1*, according to the manufacturer's specifications (Bio-
540 Rad; Hercules, CA).

541

542 ***Gene expression***

543 Total RNAs were extracted using Trizol reagent (Thermo) and the RNEasy Mini Kit (Qiagen)
544 from basophils, primary cultured mast cells, or cell lines under stated conditions and reverse-
545 transcribed using SuperScript III First-Strand Synthesis System (Thermo). ddPCR was performed
546 on a QX200 system (Bio-Rad) according to the manufacturer's instructions using custom
547 primer/probe sets for tryptase isoforms as indicated (Table S3) to quantify gene expression,

548 normalizing to *HPRT1* and *TBP* expression quantified using commercially available ddPCR
549 primer/probe sets (Bio-Rad).

550

551 ***Promoter detection assay***

552 A custom primer/probe set was designed to amplify the proximal promoter linked to alpha-tryptase
553 (Fig. S4). The forward primer corresponds to a conserved sequence present in all identified α - and
554 β -tryptases. The reverse primer hybridizes to sequences only present in alpha-tryptases. The
555 probes compete for hybridization with the C>T variant that defines $\alpha^{\text{DUP-}}$, and α^{WT} -tryptase-
556 specific sequences, respectively. To successfully amplify the large 71% GC-rich amplicon, the
557 reaction was run with 10% 1M Betaine (Sigma, Saint Louis, MO) under the following conditions:
558 95 °C 10 min, 96 °C 30 s, 66 °C 1 min (Ramp 1.5 °C), 50 cycles, 96°C 10 min.

559

560 **Exome sequencing and bioinformatic analysis for clonal variants**

561 Bone marrow aspirates were fractionated using density centrifugation into granulocyte and
562 mononuclear fractions using a double gradient 1.077 g/ml (Histopaque-1077) and 1.119 g/ml
563 (Histopaque-1119) (Sigma, St. Louis, MO). Genomic DNA was extracted and libraries were
564 prepared with the Twist Biosciences Comprehensive Exome capture kit. Deep (>300x) exome
565 sequencing was subsequently performed using the Illumina HiSeq 2500, on both the granulocyte
566 and mononuclear bone marrow fractions. Paired peripheral blood samples from the same
567 individuals were also exome sequenced consistent with clinically recommended standards (22).

568 Raw fastq files were trimmed for quality and adapter contamination using Trimmomatic
569 v0.39 (23) and mapped to the human hg38 reference genome using BWA-MEM v0.7.17
570 (<http://bio-bwa.sourceforge.net/>). PCR duplicates were marked using Samlaster v0.1.2.5 (24), and

571 GATK v4.1.9.0 was used to perform base recalibration. For the peripheral blood samples, germline
572 variation was called using GATK v4.1.9.0 and following the recommended Best Practices
573 ([https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-](https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-)
574 [discovery-SNPs-Indels-](https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-)). For granulocyte and mononuclear fractions, somatic variants were called
575 using 2 approaches. First, paired somatic calling with the PBMC sample treated as germline was
576 performed using 3 somatic detection tools: 1) MuTect2 from GATK v4.1.9.0 and following the
577 recommended Best Practices ([https://gatk.broadinstitute.org/hc/en-us/articles/360035531132--](https://gatk.broadinstitute.org/hc/en-us/articles/360035531132--How-to-Call-somatic-mutations-using-GATK4-Mutect2)
578 [How-to-Call-somatic-mutations-using-GATK4-Mutect2](https://gatk.broadinstitute.org/hc/en-us/articles/360035531132--How-to-Call-somatic-mutations-using-GATK4-Mutect2)), 2) Strelka v2.9.0 (25), and 3) mutect
579 v1.1.7 (26). In addition to paired calling, we also performed tumor-only calling with a combination
580 of MuTect2 GATK v4.1.9.0, mutect v1.1.7, and Vardict v1.7.0 (27). The latter approach of
581 performing somatic variant detection without including the paired PBMC sample was done to
582 account for the fact that somatic variants in the bone marrow granulocytes and bone marrow
583 mononuclear cells could also be present at low fractions in the PBMC sample and therefore get
584 excluded as contaminating germline variants in the paired calling. For both approaches, variants
585 were then merged across all callers and annotated with VEP v104 (28) for downstream analysis.
586 Reported variants were prioritized based upon commonly implicated genes in the development of
587 myeloid neoplasms.

588

589 **Statistical modeling and analyses**

590 To predict BST levels based upon tryptase genotypes - specifically determined by *TPSAB1*
591 replication number (including those without replications) - a log transformation was applied to the
592 BST levels because the values were not normally distributed. Additionally, we log-transformed
593 the replication number, because transforming BST back to the normal scale resulted in a model fit

594 that was approximately linear, with a near constant increase in BST for each increase in replication
595 number; this supported results from early clinical data and data generated from in silico and in
596 vitro experiments. We also added 0.5 to replication number prior to log transformation in order to
597 include individuals without *TPSAB1* replications in the model, thus allowing us to also determine
598 a cutoff for individuals without HαT. Thus, we used a linear regression model to predict log(BST)
599 from log(replication number plus 0.5), and from that, created an upper one-sided 99.5% prediction
600 interval. The 99.5% threshold was chosen as it provides high specificity where very few
601 individuals will be identified as false positives and correlated well with the predicted prevalence
602 of elevated BST in the absence of HαT based upon available population data (10). The model was
603 developed using R version 3.6.3.

604 For clinical and experimental data, Mann-Whitney, Kruskal-Wallis, and paired and
605 unpaired two-tailed t-tests, were used where appropriate to test significance of differences,
606 prevalence, or deviation using Prism (GraphPad).

607

608 **Basal Serum Tryptase (BST) calculator code**

609 We developed an online calculator the **B**asal **S**erum **T**ryptase **C**linical cut-off **A**ssigned by **L**ocus
610 **C**opy number of **U**TR-**L**inked element and **A**ssociated **T***TPSAB1* **E**ncoded **R**eplication (BST
611 CALCULATOR) with Shiny R framework: <https://bst-calculater.niaid.nih.gov/>. The code is
612 available at: <https://github.com/niaid/BST-calculater>

613

614 **URLs**

615 Burrows-Wheeler Aligner <http://bio-bwa.sourceforge.net/>

616 Picard, <http://broadinstitute.github.io/picard/>

617 Plink, <http://pngu.mgh.harvard.edu/purcell/plink/>

618 Genome Analysis Toolkit (GATK), <http://www.broadinstitute.org/gatk/>

619 SAMtools, <http://www.htslib.org/>

620

621 Accession Codes

622 Sequencing data were obtained from the National Center for Biotechnology Information Sequence

623 Read Archive for the following BioProjects: hereditary alpha-tryptasemia: **PRJNA342304**;

624 individuals not selected by tryptase genotype: **PRJNA208369 PRJNA219425 PRJNA232669**

625 **PRJNA252605 PRJNA253059 PRJNA254943 PRJNA257389 PRJNA258216**

626 **PRJNA261011 PRJNA261251 PRJNA263242 PRJNA266512 PRJNA266572**

627 **PRJNA270371 PRJNA271942 PRJNA274028 PRJNA274360 PRJNA275801**

628 **PRJNA279249 PRJNA280990 PRJNA283839 PRJNA289905 PRJNA291619**

629 **PRJNA292690 PRJNA293555 PRJNA296379 PRJNA301173 PRJNA301364**

630 **PRJNA310988 PRJNA315611 PRJNA317535 PRJNA318253 PRJNA319220**

631 **PRJNA326113 PRJNA327986 PRJNA330840 PRJNA340161 PRJNA342177**

632 **PRJNA343985 PRJNA354367 PRJNA358081 PRJNA369563 PRJNA369684**

633 **PRJNA373887 PRJNA376200 PRJNA377555 PRJNA378385 PRJNA384963**

634 **PRJNA388978 PRJNA389466 PRJNA392116 PRJNA395367 PRJNA395589**

635 **PRJNA399103 PRJNA400331 PRJNA407731 PRJNA415746 PRJNA428940**

636 **PRJNA214592 PRJNA229548 PRJNA278364 PRJNA278767 PRJNA263397**

637

638 Referenced GenBank Accession Codes

639 GRCh37/hg19 NCBI assembly: **GCF_000001405.13**; tryptase locus assemblies: **AC226137.3**

640 **AC120498.2 CTD-2503P16 CHM13 AL031704.24 AE006466.1 CHM1 AC240106.3**

641 **AC238650.2 AC213746.1 GM24385 NA12878 AF098328.1**

642

643 **Author Contributions:** J.J.L. designed the study. J.J.L., S.L., C.N., J.B., T.D., M.C.C., H.D.K.,
644 J.S., K.M.H., E.F., P.K., J.D.M., V.S., D.E., P.B., R.Z., P.K., and D.D.M, recruited and enrolled
645 study participants. J.J.L., I.T., J.H., J.H., J.M., C.S.H., B.B., M.P., N.N.T., R.E.H., S.A.M., G.H.C.,
646 and J.L. performed and/or supervised the bioinformatics analyses. J.J.L. designed and J.C., J.K.,
647 and Y.L. performed ddPCR assays. I.M. performed hematopathologic analyses. J.J.L., J.C., J.K.,
648 D.D.M., Y.B., Y.Y., Y.H.P., Y.L., M.P.O., and A.P.H. designed, supervised, and /or performed
649 the functional and genetic studies and sequencing, respectively. A.M. and E.H.B. developed the
650 prediction interval model. J.J.L. prepared the draft manuscript. All authors contributed to
651 discussion of the results and to manuscript preparation.

652

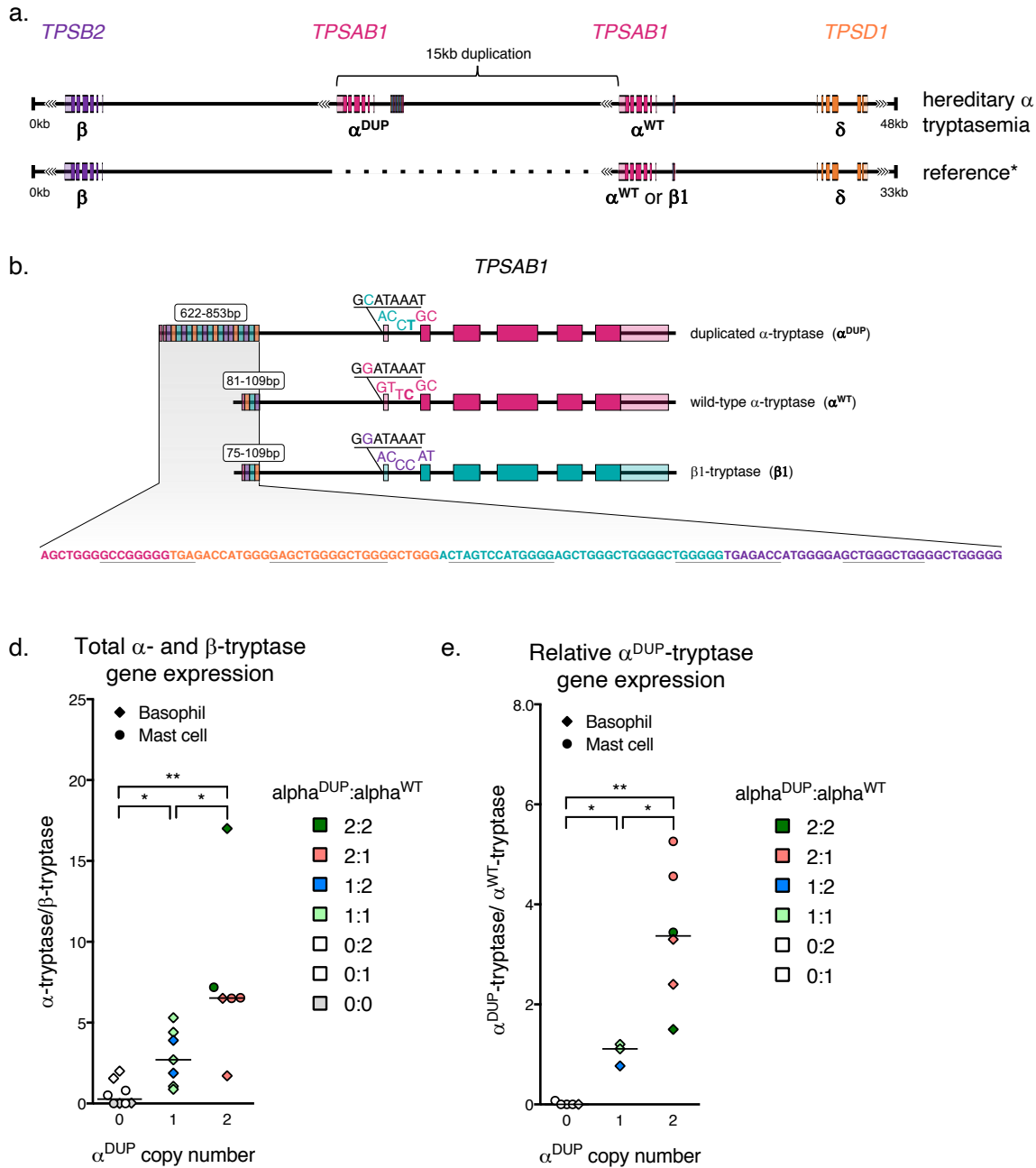
653 **Acknowledgements:** The investigators thank the patients, their families, and healthy volunteers
654 who contributed to this research. We also thank Adam Phillippy for sharing his expertise on
655 complex genetic structural rearrangement.

References

- 656
657
658 1. Caughey GH. Tryptase genetics and anaphylaxis. *J Allergy Clin Immunol.*
659 2006;117(6):1411-4.
660 2. Miller JS, Westin EH, Schwartz LB. Cloning and characterization of complementary
661 DNA for human tryptase. *J Clin Invest.* 1989;84(4):1188-95.
662 3. Miller JS, Moxley G, Schwartz LB. Cloning and characterization of a second
663 complementary DNA for human tryptase. *J Clin Invest.* 1990;86(3):864-70.
664 4. Glover SC, Carter MC, Korosec P, Bonadonna P, Schwartz LB, Milner JD, et al. Clinical
665 relevance of inherited genetic differences in human tryptases: Hereditary alpha-tryptasemia and
666 beyond. *Ann Allergy Asthma Immunol.* 2021.
667 5. Lyons JJ, Yu X, Hughes JD, Le QT, Jamil A, Bai Y, et al. Elevated basal serum tryptase
668 identifies a multisystem disorder associated with increased TPSAB1 copy number. *Nat Genet.*
669 2016;48(12):1564-9.
670 6. Chollet MB, Akin C. Hereditary alpha tryptasemia is not associated with specific clinical
671 phenotypes. *J Allergy Clin Immunol.* 2021.
672 7. Greiner G, Sprinzl B, Gorska A, Ratzinger F, Gurbisz M, Witzeneder N, et al. Hereditary
673 alpha tryptasemia is a valid genetic biomarker for severe mediator-related symptoms in
674 mastocytosis. *Blood.* 2020.
675 8. Lyons JJ, Chovanec J, O'Connell MP, Liu Y, Selb J, Zanotti R, et al. Heritable risk for
676 severe anaphylaxis associated with increased alpha-tryptase-encoding germline copy number at
677 TPSAB1. *J Allergy Clin Immunol.* 2020.
678 9. Robey RC, Wilcock A, Bonin H, Beaman G, Myers B, Grattan C, et al. Hereditary
679 Alpha-Tryptasemia: UK Prevalence and Variability in Disease Expression. *J Allergy Clin*
680 *Immunol Pract.* 2020.
681 10. Lyons JJ. Inherited and acquired determinants of serum tryptase levels in humans. *Ann*
682 *Allergy Asthma Immunol.* 2021;In press.
683 11. Lyons JJ. Hereditary Alpha Tryptasemia: Genotyping and Associated Clinical Features.
684 *Immunol Allergy Clin North Am.* 2018;38(3):483-95.
685 12. Schwartz LB, Metcalfe DD, Miller JS, Earl H, Sullivan T. Tryptase levels as an indicator
686 of mast-cell activation in systemic anaphylaxis and mastocytosis. *N Engl J Med.*
687 1987;316(26):1622-6.
688 13. Lyons JJ, Schwartz LB. Clinical Approach to a Patient with Elevated Serum Tryptase:
689 Implications of Acute Versus Basally Elevated Levels. 1 ed. Akin C, editor: Springer, Cham;
690 2020.
691 14. Valent P, Horny HP, Escribano L, Longley BJ, Li CY, Schwartz LB, et al. Diagnostic
692 criteria and classification of mastocytosis: a consensus proposal. *Leuk Res.* 2001;25(7):603-25.
693 15. Luskin KT, White AA, Lyons JJ. The Genetic Basis and Clinical Impact of Hereditary
694 Alpha-Tryptasemia. *J Allergy Clin Immunol Pract.* 2021.
695 16. Trivedi NN, Tamraz B, Chu C, Kwok PY, Caughey GH. Human subjects are protected
696 from mast cell tryptase deficiency despite frequent inheritance of loss-of-function mutations. *J*
697 *Allergy Clin Immunol.* 2009;124(5):1099-105 e1-4.
698 17. Mateja A, Wang Q, Chovanec J, Kim J, Wilson KJ, Schwartz LB, et al. Defining baseline
699 variability of serum tryptase levels improves accuracy in identifying anaphylaxis. *J Allergy Clin*
700 *Immunol.* 2021;In revision.

- 701 18. Huang R, Abrink M, Gobl AE, Nilsson G, Aveskogh M, Larsson LG, et al. Expression of
702 a mast cell tryptase in the human monocytic cell lines U-937 and Mono Mac 6. *Scand J*
703 *Immunol.* 1993;38(4):359-67.
- 704 19. Soto D, Malmsten C, Blount JL, Muilenburg DJ, Caughey GH. Genetic deficiency of
705 human mast cell alpha-tryptase. *Clin Exp Allergy.* 2002;32(7):1000-6.
- 706 20. Lyons JJ, Stotz SC, Chovanec J, Liu Y, Lewis KL, Nelson C, et al. A common haplotype
707 containing functional CACNA1H variants is frequently coinherited with increased TPSAB1
708 copy number. *Genet Med.* 2018;20(5):503-12.
- 709 21. Min HK, Moxley G, Neale MC, Schwartz LB. Effect of sex and haplotype on plasma
710 tryptase levels in healthy adults. *J Allergy Clin Immunol.* 2004;114(1):48-51.
- 711 22. Waters AM, Park HJ, Weskamp AL, Mateja A, Kachur ME, Lyons JJ, et al. Elevated
712 basal serum tryptase: disease distribution and variability in a regional health system. *J Allergy*
713 *Clin Immunol Pract.* 2022.
- 714
- 715

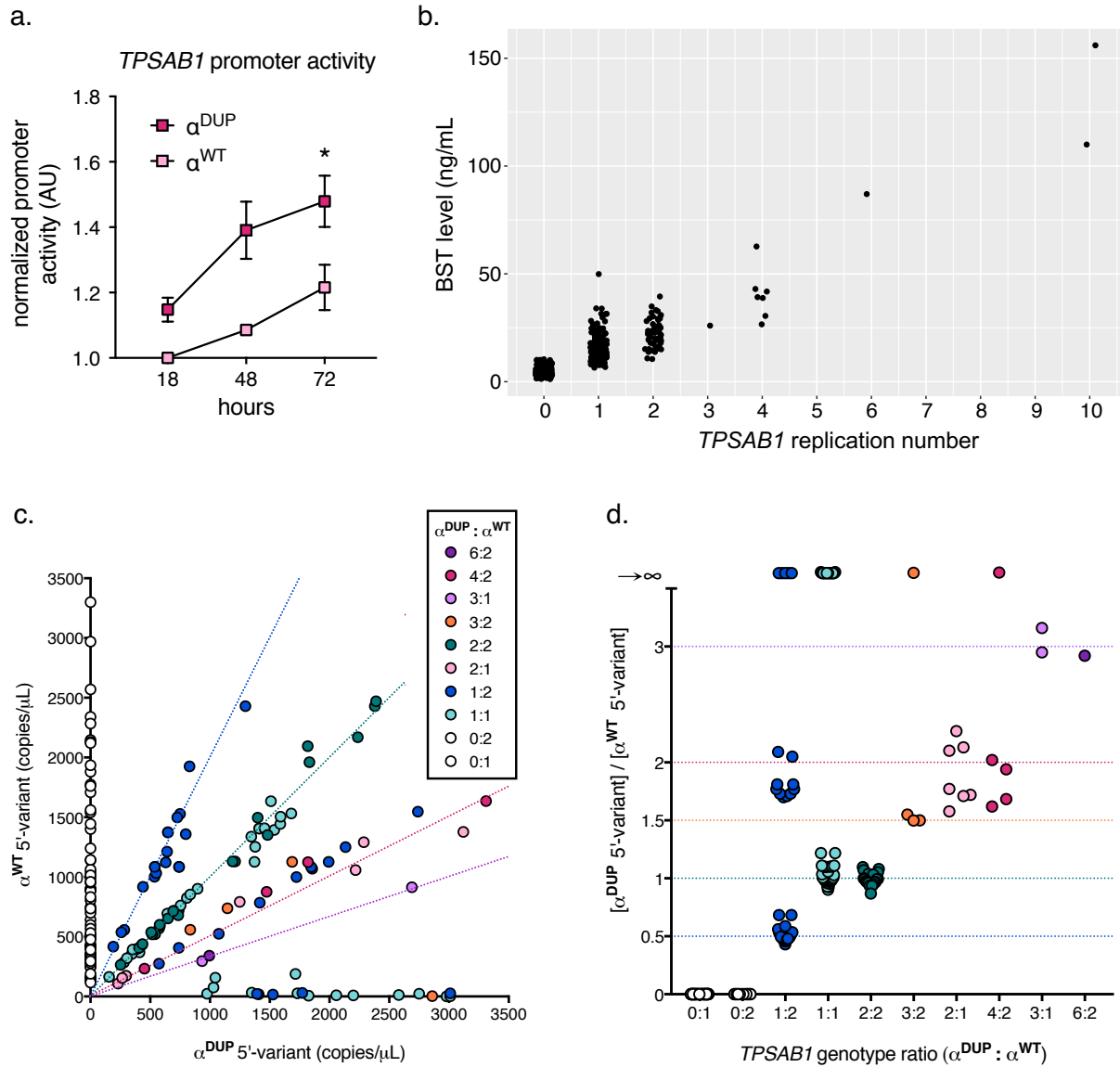
716 FIGURES



717

718 **Figure 1. *TPSAB1* duplications occur at 16p13.3 and are linked to an expanded promoter**
 719 **associated with α^{DUP} -tryptase over-expression. (a)** Location and orientation of the 15-kB
 720 tandem duplication of *TPSAB1*. **(b)** Alignment of duplicated alpha- (α^{DUP}), wild-type alpha (α^{WT}),
 721 and beta1-tryptases with unique 5'-variants and size of promoter repeat regions indicated; repeat
 722 motif within promoter regions **(c)**. Relative total α - to β -tryptase gene expression **(d)** and α^{DUP} -
 723 relative to α^{WT} -tryptase gene expression **(e)** in ex-vivo basophils and cultured primary mast cells
 724 (right). * $P < 0.05$; ** $P < 0.005$.

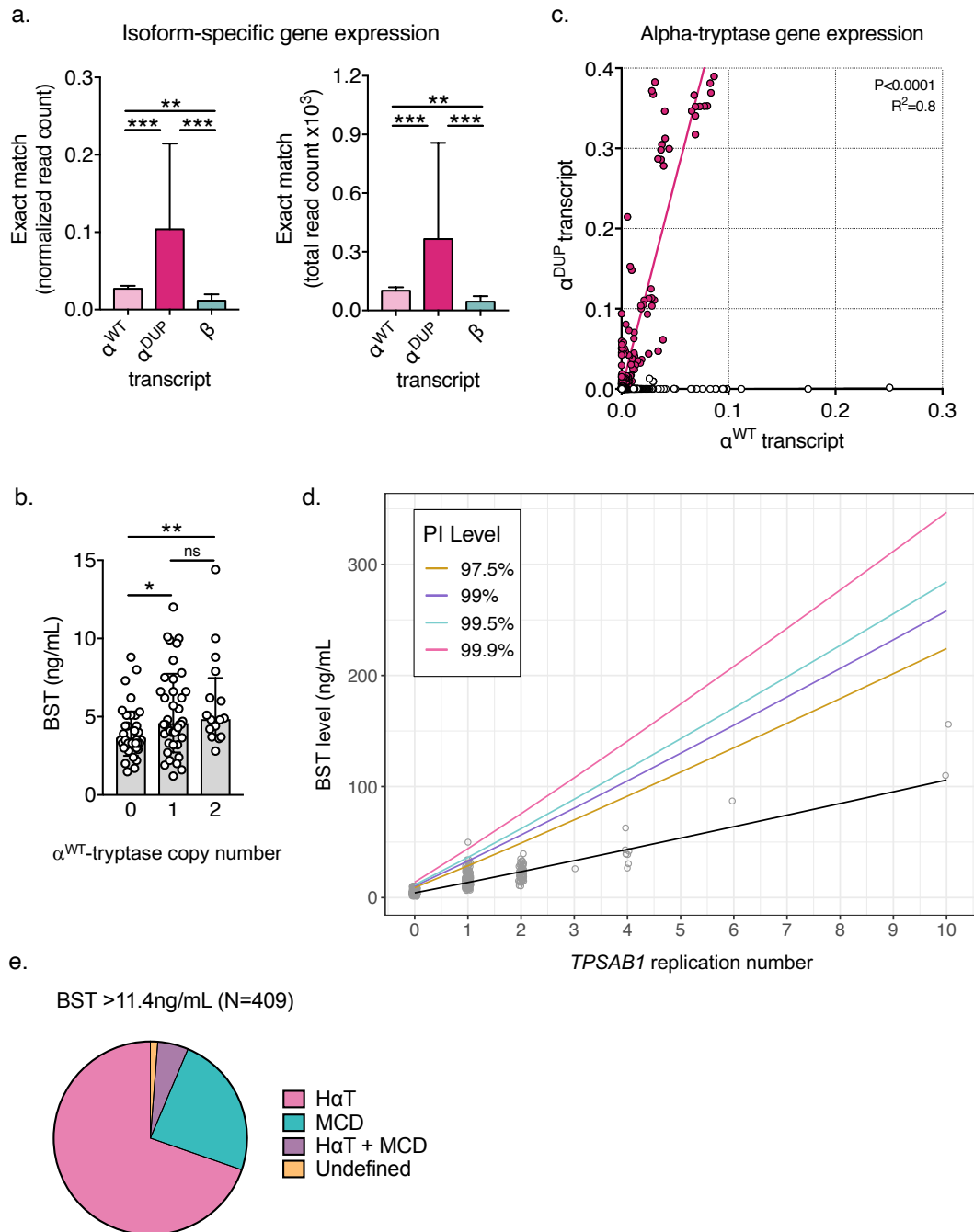
725



726

727 **Figure 2. The expanded promoter at duplicated *TPSAB1* is conserved and has increased**
 728 **basal activity. (a)** Normalized fluorescent-reporter activity in unstimulated MonoMac-6 cells
 729 transfected with promoters cloned from endogenous (α^{WT}) or replicated (α^{DUP}) *TPSAB1* loci. **(b)**
 730 Basal serum tryptase (BST) from individuals grouped by *TPSAB1* replication number. Relative
 731 allelic frequency **(c)** and ratios **(d)** of α^{WT} - (y-axis) and α^{DUP} -tryptase (x-axis) associated 5'-
 732 variants determined by droplet digital PCR; dashed lines indicate predicted $\alpha^{WT}:\alpha^{DUP}$ ratios by
 733 genotype. *P<0.05

734



735 **Figure 3. Modeling serum tryptase levels based upon genotype improves clinical utility.** (a)
 736 Normalized (left) and total (right) read counts for reads aligning exactly to the 39-bp consensus
 737 sequences that identify β -, α^{WT} - and α^{DUP} -tryptase. (b) BST levels among individuals with
 738 conserved 4n tryptase copy number (combined from *TPSAB1* and *TPSB2*). (c) Regression analysis
 739 of relative expression levels of α^{DUP} (y-axis) and α^{WT} (x-axis) tryptase transcripts. (d) Prediction
 740 intervals (PI) for BST levels based upon *TPSAB1* replication number. (e) Prevalence of HaT,
 741 clonal mast cell disease (MCD), and those without either, among individuals referred with BST
 742 levels above the predicted upper limit of normal (>11.4 ng/mL). * $P < 0.01$; ** $P < 0.005$;
 743 *** $P < 0.0001$

744 TABLES

Table 1. Assemblies of the human tryptase locus at 16p13.3.

<i>TPSB2</i>			<i>TPSAB1^{REP}</i>			<i>TPSAB1</i>			<i>TPSD1</i>		Corresponding Assembly
haplotype	orientation	isoform	haplotype	orientation	isoform	haplotype	orientation	isoform	haplotype	orientation	
CACCT	REV	beta3	-	-	-	GACCC	FWD	beta1	GGTTC	FWD	GRCh37/hg19
-	-	-	-	-	-	GACCC	REV	beta1	GGTTC	FWD	AC226137.3
CACCT	REV	beta3	-	-	-	GACCC	FWD	beta1	CGTTC	FWD	AC120498.2
CACCT	REV	beta3	-	-	-	GACCC	REV	beta1	CGTTC	FWD	CTD-2503P16
CACCT	REV	beta3	-	-	-	GACCC	REV	beta1	CGTTC	FWD	CHM13
CACCT	REV	beta3	-	-	-	CACCT	REV	beta3	GGTTC	FWD	AL031704.24
CACCT	REV	beta3	-	-	-	CACCT	REV	beta3	GGTTC	FWD	AE006466.1
GACCT	REV	beta2	-	-	-	GGTTC	FWD	alpha	GGTTC	FWD	CHM1
GACCT	REV	beta2	-	-	-	GGTTC	FWD	alpha	GGTTC	FWD	AC240106.3
GACCT	REV	beta2	-	-	-	GGTTC	FWD	alpha	GGTTC	FWD	AC238650.2
GACCT	REV	beta2	-	-	-	GGTTC	FWD	alpha	GGTTC	FWD	AC213746.1
GACCT	REV	beta2	-	-	-	GGTTC	REV	alpha	GGTTG	FWD	GM24385
GACCT	REV	beta2	-	-	-	GGTTC	REV	alpha	GGTTC	FWD	NA12878
GACCC	REV	beta1	CACCT	REV	alpha	GGTTC	REV	alpha	GGTTC	FWD	HαT BAC
-	-	-	-	-	-	GGTTC	-	alpha	-	-	AF098328.1

745 Bold – assembly of the locus from an individual with hereditary alpha-tryptasemia (HαT)

746

Table 2. Haplotypes identified in linkage with allelic tryptase genotypes.

<i>TPSB2</i>		<i>TPSAB1^{DUP}</i>		<i>TPSAB1</i>		<i>TPSD1</i>	Allele count	Allelic genotype	Haplotype Frequency (by genotype)
haplotype	isoform	haplotype	isoform	haplotype	isoform	haplotype			
CACCT	beta3	-	-	GACCC	beta1	GGTTC	172	β/β	52%
CACCT	beta3	-	-	GACCC	beta1	CGTTC	152	β/β	46%
GACCT	beta2	-	-	GACCC	beta1	GGTTC	7	β/β	2%
GACCT	beta2	-	-	GGTTC	alpha	GGTTC	65	α/β	96%
GACCT	beta2	-	-	GGTTC	alpha	CGTTC	3	α/β	4%
GACCC	beta1	CACCT	alpha	GGTTC	alpha	GGTTC	15	αα/β	63%
GACCC	beta1	CACCT	alpha	---- T	alpha	GGTTC	5	αα/β	21%
GACCC	beta1	CACCT	alpha	GGTT (-6delC)*	alpha	GGTTC	2	αα/β	8%
GACCC	beta1	CACCT	alpha	GGTTT	alpha	GGTTC	2	αα/β	8%

*A single cytosine (C) base deletion was identified 6 bases prior (-6) to the terminal C of the haplotype

747
748

Table 3. Measured and predicted BST levels based upon *TPSAB1* replication number encoding α -tryptase.

<i>TPSAB1</i> replication number	Primary data			Modeled reference values	
	N	Median BST	SD BST	Fit	Upper 99.5% PI
0	204	4.15	2.12	4.2831	11.3879
1	247	13.60	5.55	13.6363	36.2216
2	52	22.45	6.40	23.3642	62.1517
3	1	26.00	NA	33.3110	88.7541
4	6	40.50	12.61	43.4148	115.8535
5	-	-	-	53.6420	143.3508
6	1	87.00	NA	63.9708	171.1816
7	-	-	-	74.3862	199.3007
8	-	-	-	84.8774	227.6741
9	-	-	-	95.4357	256.2756
10	2	133.00	32.53	106.0544	285.0841

BST - basal serum tryptase; SD - standard deviation; PI - prediction interval

Table 4. Individuals with MCD and H α T.

Diagnosis	Age	Gender	Tryptase genotype		Current WHO-defined criteria for SM diagnosis							
			α	β	MC aggregates*	Spindle [†]	<i>KIT</i> p.D816V [†]	CD2 or CD25 [†]	BST	>20ng/mL [†]	Above 99.5% PI	New Diagnosis [‡]
ISM	66-70	M	3	2	yes	yes	positive	positive	236	yes	yes	no
ISM	20-25	M	4	2	yes	yes	positive	positive	107	yes	yes	no
ISM	66-70	F	3	2	yes	yes	positive	positive	88.7	yes	yes	no
ISM	66-70	F	3	2	yes	yes	negative	positive	70	yes	yes	no
ISM	10-15	F	2	3	yes	yes	positive	positive	60.1	yes	yes	no
MMAS	70-75	F	3	2	no	no	negative	positive	52.2	yes	yes	no
ISM	50-55	M	3	2	no	yes	positive	positive	51	yes	yes	no
ISM	70-75	F	3	2	no	yes	positive	negative	42	yes	yes	no
MMAS	66-70	F	2	3	no	no	positive	negative	38.8	yes	yes	no
ISM (WD)	30-35	F	3	2	yes	no	negative	negative	37.9	yes	yes	no
ISM	10-15	M	3	2	no	yes	positive	positive	37.6	yes	yes	no
MMAS	20-25	M	6	2	no	no	positive	negative	37	yes	yes	no
DCM	6-10	M	3	2	NA	NA	negative [§]	NA	32.6	yes	no	no
ISM	56-60	F	2	3	no	yes	positive	positive	32.2	yes	no	no
MMAS	50-55	F	3	2	no	yes	negative	negative	27.3	yes	no	no
MMAS	60-65	M	3	2	no	no	positive	negative	26.8	yes	no	no
<i>ISM (WD)</i>	<i>26-30</i>	<i>F</i>	<i>2</i>	<i>3</i>	<i>yes</i>	<i>no</i>	<i>negative</i>	<i>negative</i>	<i>26.8</i>	<i>yes</i>	<i>no</i>	<i>yes</i>
ISM	30-35	F	3	2	no	yes	positive	positive	23.7	yes	no	no
MMAS	60-65	M	2	3	no	no	positive	negative	23.1	yes	no	no
MMAS	56-60	M	3	2	no	no	positive	negative	20.3	yes	no	no
MMAS	66-70	F	2	3	no	no	positive	negative	18.8	no	no	no

ISM- indolent systemic mastocytosis; WD - well-differentiated; MMAS - monoclonal mast cell activation syndrome; DCM - diffuse cutaneous mastocytosis; *major and [†]minor criteria for the diagnosis of SM; [‡]a new diagnosis being assigned based upon application of the 99.5% threshold for BST, rather than >20ng/mL; [§]in peripheral blood only. The individual with a change in diagnosis based upon genotype-defined BST reference values is indicated in bold italic font.