

1 **Human IgE producing B cells have a unique transcriptional program and generate high**
2 **affinity, allergen-specific antibodies**

4 Derek Croote¹, Spyros Darmanis⁶, Kari C. Nadeau^{3,4,5}, Stephen R. Quake^{1,2,6*}

5
6 Departments of Bioengineering¹, Applied Physics², Medicine³, and Pediatrics⁴, Stanford
7 University, Stanford, CA 94305, USA

8 ⁵Sean N. Parker Center for Allergy and Asthma Research at Stanford University

9 ⁶Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

10 *Corresponding author. E-mail: quake@stanford.edu

11 **Abstract**

12 IgE antibodies provide defense against helminth infections, but can also cause life-threatening
13 allergic reactions. Despite their importance to human health, these antibodies and the cells that
14 produce them remain enigmatic due to their scarcity in humans; much of our knowledge of their
15 properties is derived from model organisms. Here we describe the isolation of IgE producing B
16 cells from the blood of individuals with food allergies, followed by a detailed study of their
17 properties by single cell RNA sequencing (scRNA-seq). We discovered that IgE B cells are
18 deficient in membrane immunoglobulin expression and that the IgE plasmablast state is more
19 immature than that of other antibody producing cells. Through recombinant expression of
20 monoclonal antibodies derived from single cells, we identified IgE antibodies which had
21 unexpected cross-reactive specificity for major peanut allergens Ara h 2 and Ara h 3; not only
22 are these among the highest affinity native human antibodies discovered to date, they represent
23 a surprising example of convergent evolution in unrelated individuals who independently
24 evolved nearly identical antibodies. Finally, we discovered that splicing within B cells of all
25 isotypes reveals polarized germline transcription of the IgE, but not IgG4, isotype as well as
26 several examples of biallelic expression of germline transcripts. Our results offer insights into
27 IgE B cell transcriptomics, clonality and regulation, provide a striking example of adaptive
28 immune convergence, and offer an approach for accelerating mechanistic disease
29 understanding by characterizing a rare B cell population underlying IgE-mediated disease at
30 single cell resolution.

31 **Introduction**

32 The IgE antibody class is the least abundant of all isotypes in humans and plays an important
33 role in host defense against parasitic worm infections (1), but it can also become misdirected
34 towards otherwise harmless antigens. Food allergies are one example of this misdirection,
35 where the recognition of allergenic food proteins by IgE antibodies can lead to symptoms
36 ranging from urticaria to potentially fatal anaphylaxis. Despite this central role in immunity and
37 allergic disease, human IgE antibodies remain poorly characterized due to their scarcity (2).
38 Bulk epitope mapping experiments have revealed that IgE antibodies are polyclonal and
39 epitopes are heterogeneous (3); however, individuals with the same allergy tend to recognize a
40 core set of one or a few allergenic proteins (4). Recent studies applying bulk fluorescence

45 activated cell sorting (FACS) immunophenotyping (5, 6) and immune repertoire deep
46 sequencing (7) have inferred IgE B cell characteristics and origins, while studies performing
47 peanut allergen specific single cell sorting (8, 9) have described clonal families to which IgE
48 antibodies belong. However, none have successfully isolated single IgE producing cells or the
49 paired heavy and light chain sequences that comprise individual IgE antibodies, leaving
50 unanswered questions as to the functional properties of such antibodies, the transcriptional
51 programs of these cells, and the degree to which any of these features are shared across
52 individuals. Similarly, there is a lack of knowledge, but growing interest, surrounding the IgG4
53 isotype due to its potential to compete with IgE for allergen and thereby contribute to the
54 reduced clinical allergen reactivity that accompanies immunotherapy and early allergen
55 exposure (10). Here we report the first successful isolation and transcriptomic characterization
56 of single IgE and IgG4 producing B cells from humans. We combined single cell RNA
57 sequencing (scRNA-seq) with functional antibody assays to elucidate mechanisms underlying
58 the regulation of IgE and to discover high affinity peanut-specific antibodies.
59

60 **Characterization of single B cells from peripheral blood**

61

62 We performed scRNA-seq on B cells isolated from the peripheral blood of food allergic
63 individuals, which enabled us to characterize each cell's gene expression, splice variants, and
64 heavy and light chain antibody sequences (Fig. 1A). Fresh peripheral blood from six peanut
65 allergic individuals was first separated into plasma and cellular fractions; plasma was stored and
66 later used for allergen-specific IgE concentration measurements (fig. S1), while the cellular
67 fraction was enriched for B cells prior to FACS (see Materials and Methods). CD19+ B cells of
68 all isotypes were sorted exclusively based on immunoglobulin surface expression, but with an
69 emphasis on maximizing IgE B cell capture (fig. S2). Because cellular isotype identity was
70 determined post-hoc from scRNA-seq, we were able to sacrifice specificity and capture cells
71 with high sensitivity. This approach makes the prospect of IgE B cell capture accessible for
72 many laboratories without stringent requirements on FACS gate purity or the need for complex,
73 many-color gating schemes.
74

75 Single cells were sorted into 96 well plates, processed using a modified version of the Smart-
76 seq2 protocol (11), and sequenced on an Illumina NextSeq 500 with 2x150 bp reads to an
77 average depth of 1-2 million reads per cell (fig. S3). Sequencing reads were independently
78 aligned and assembled to produce a gene expression count table and reconstruct antibody
79 heavy and light chains, respectively (see Materials and Methods). Using STAR (12) for
80 alignment also facilitated the assessment of splicing within single cells. Cells were stringently
81 filtered to remove those of low quality, putative basophils, and those lacking a single productive
82 heavy and light chain, yielding a total of 973 cells for further analysis. The isotype identity of
83 each cell was determined by its productive heavy chain assembly, which avoids
84 misclassification of isotype based on FACS immunoglobulin surface staining (fig. S2B), a
85 problem which is especially pervasive for IgE B cells due to CD23, the "low-affinity" IgE receptor
86 that captures IgE on the surface of non-IgE B cells (6).
87

88 Principal component analysis of normalized gene expression following batch effect correction
89 (fig. S3 and Materials and Methods) separated cells into two distinct clusters identifiable as
90 plasmablasts (PBs) and naïve / memory B cells (Fig. 1B-C). PBs expressed the triad of
91 transcription factors BLIMP1 (PRDM1), XBP1, and IRF4 that drive plasma cell differentiation
92 (13), as well as genes associated with antibody secretion (fig. S4), while naïve and memory
93 cells expressed the canonical mature B cell surface marker CD20 (MS4A1), as well as
94 transcription factor IRF8, which antagonizes the PB fate and instead promotes a germinal
95 center response (14). Additional data corroborated this cell subtype assignment; PBs had
96 greater FACS forward and side scatter in agreement with their larger size and increased
97 granularity, PB cDNA concentrations were higher following preamplification, and PBs expressed
98 more antibody heavy and light chain transcripts (fig. S4).

99

100 We assessed the distribution of isotypes within each B cell subtype and found that, in stark
101 contrast to other isotypes, circulating IgE B cells overwhelmingly belonged to the PB subtype
102 (Fig 1D, fig. S5A). This discovery is consistent with observations of preferential differentiation of
103 IgE B cells into PBs in mice (15). Subtype proportions for other isotypes followed expectations:
104 IgM B cells, which are primarily naïve, had the lowest PB percentage, while IgA B cells had the
105 highest in accordance with their secretory role in maintaining mucosal homeostasis.
106 Interestingly, we found that the number of circulating IgE B cells for each individual correlated
107 with total plasma IgE levels (fig S1C); a similar phenomenon has been noted in atopic
108 individuals and individuals with hyper-IgE syndrome (16).

109

110 By clustering antibodies into clonal families (CFs) we were able to observe elements of classical
111 germinal center phenomena such as somatic hypermutation, class switching, and fate
112 determination in our data. Using a standard immune repertoire sequencing approach (17), all
113 antibody heavy chain sequences were first divided by V and J genes and were clustered if their
114 amino acid CDR3 sequences shared at least 75% similarity. Only 49 heavy chains formed CFs
115 with multiple members, although this was not surprising given the vast diversity of potential
116 immunoglobulin gene rearrangements (fig. S5B). Within multi-member CFs, light chains were
117 highly similar (fig. S5D), while overall, multi-member CFs were diverse (Fig. 1E); they contained
118 between two and six sequences, had variable isotype membership, and had a comprehensive
119 distribution of mutational frequency. CFs were specific to an individual, with the exception of one
120 CF (CF1) that contained six heavily mutated IgE PBs: three each from individuals PA12 and
121 PA13, as discussed in depth later. Four CFs illustrated the two possible B cell differentiation
122 pathways in that they contained both PBs and memory B cells. Other CFs contained cells
123 belonging to multiple isotypes, with one of particular interest (CF3), discussed later, that
124 contained an IgE PB and an IgG4 PB. Interestingly, we found that in contrast to other isotypes,
125 IgE and IgG4 were surprisingly clonal as over 20% of IgE and IgG4 cells belonged to such
126 multi-member CFs (fig. S5C).

127

128 Across all individuals, the 89 IgE antibodies we found varied widely in gene usage and mutation
129 frequency (Fig. 2A). They also varied in heavy and light chain CDR3 lengths (fig. S6A). There
130 was moderate correlation between the mutation frequency of heavy and light chains within
131 single cells (fig. S6B), with evidence of selection via an enrichment of replacement mutations

132 relative to silent mutations in the heavy chain CDR1 and CDR2 that was absent in framework
133 (FWR) regions. Light chains were similarly enriched for replacement mutations in the CDR1
134 and, to a lesser degree, FWR2 (fig. S6C). Compared to other isotypes, IgE B cells had a similar
135 distribution of heavy chain mutation frequency, relative utilization of the lambda versus kappa
136 light chains, and heavy chain V and J gene usage (fig. S6D-F).

137

138 **IgE B cells possess a unique transcriptional program**

139

140 To elucidate B cell intrinsic factors affecting PB activation, survival, and differentiation, we
141 assessed genes differentially expressed between IgE PBs and PBs of other isotypes (Fig. 2B).
142 A host of MHC genes were robustly upregulated in IgE PBs, suggesting a more immature
143 transcriptional program given the established loss of MHC-II during the maturation of PBs to
144 plasma cells (18–20). FCER2 (CD23), the “low-affinity” IgE receptor was also highly
145 upregulated, although its precise role within IgE PBs is unclear; autoinhibition of IgE production
146 could result from membrane CD23-mediated co-ligation of membrane IgE (mlgE) and CD21
147 (21). Alternatively, IgE production could be upregulated by soluble CD23 (22), which is
148 produced following cleavage by ADAM10 (23), a disintegrin and metalloproteinase domain-
149 containing protein that we find is co-expressed in a subset of IgE PBs. LAPTM5, a negative
150 regulator of B cell activation, BCR expression, and antibody production (24), was also
151 upregulated, while CSF2RB, which encodes the common beta chain of the IL-3 and IL-5
152 receptors, was downregulated, potentially indicating weakened IL-3- and IL-5-mediated terminal
153 differentiation capacity (25, 26). Additional downregulated genes included galectin 1 (LGALS1),
154 which supports plasma cell survival (27) and the S100 proteins S100A4, S100A6, and
155 S100A10, which may indicate reduced proliferative and survival signaling (28, 29). One of the
156 most significantly downregulated genes in IgE PBs was spleen associated tyrosine kinase
157 (SYK), which plays an essential role in BCR signal transduction (30) and is necessary for naïve
158 B cell differentiation into plasma cells and for memory B cell survival (31). Taken together, this
159 gene expression program shows that the IgE PB cell state is immature relative to other PBs with
160 weakened activation, proliferation, and survival capacity. It also provides a potential
161 transcriptomic mechanism for the hypothesized short-lived IgE PB phenotype described in
162 mouse models of allergy (15, 32).

163

164 We found human IgE B cells belonging to the naïve / memory subset were deficient in
165 immunoglobulin heavy chain membrane exon splicing compared to other common isotypes.
166 Furthermore, membrane exon splicing was detected at low levels in non-IgE PBs, but not in IgE
167 PBs (Fig. 2C-D). In fact, the absence of mlgE splicing rendered us unable to assess the relative
168 utilization of the two splice variants of mlgE known to have distinct signaling characteristics (33,
169 34). The lack of mature mlgE transcripts could be explained by poor processing of pre-mRNA
170 (35) and is consistent with low IgE surface protein we measured by FACS; indeed, mlgE
171 surface protein levels on true IgE B cells did not exceed those of some non-IgE B cells
172 presumably displaying surface IgE as a result of CD23-mediated capture (fig. S2B). These
173 results suggest that the scarcity of circulating memory IgE B cells *in vivo* could result from
174 impaired membrane IgE expression that compromises IgE B cell entry into the memory
175 compartment and/or memory B cell survival. Murine studies support such a hypothesis, having

176 shown IgE responses are reduced by removal or modification of mlgE domains, but augmented
177 by the exchange of these domains for those of IgG1, thereby suggesting that poor mlgE
178 expression and signaling acts to restrict IgE levels (36, 37).

179

180 **Characterization of peanut-specific IgE and IgG4 antibodies**

181

182 Surprisingly, our clonal analysis produced one CF of cells belonging to multiple individuals
183 (CF1, Fig. 1E), which contained three IgE PBs from individual PA12 and three IgE PBs from
184 individual PA13. The antibodies produced by these six cells were highly similar (Fig. 3A, fig.
185 S7A-B) as all utilized the IGHV3-30*18 and IGHJ6*02 heavy chain genes as well as the IGKV3-
186 20*01 and IGKJ2*01 light chain genes, with pairwise CDR3 amino acid sequence identity
187 ranging from 65% to 94% for the heavy chain and 70% to 100% for the light chain. These
188 antibodies were also highly mutated and enriched in replacement mutations within the
189 complementarity determining regions of both chains (fig. S7C). In fact, compared to all other
190 class switched antibodies, these were amongst the most mutated: the heavy chains were in the
191 76th percentile or above for mutation frequency, while all of the light chains were in the 96th
192 percentile or above (fig. S7D).

193

194 We recombinantly expressed the six IgE antibodies belonging to this convergent clonal family in
195 order to assess whether they bind the natural forms of the major allergenic peanut (*Arachis*
196 *hypogaea*) proteins Ara h 1, Ara h 2, or Ara h 3. Of all characterized peanut allergens, Ara h 2 is
197 the most commonly recognized by allergic individuals and is the most clinically relevant both in
198 terms of immunological response (38) and discriminating allergic status (39, 40). Using an
199 indirect ELISA as a qualitative screen for binding, we found that, surprisingly, all six antibodies
200 were cross-reactive; they bound strongly to Ara h 2, moderately to Ara h 3, and very weakly to
201 Ara h 1 (Fig. 3B). We then used biolayer interferometry to determine dissociation constants of
202 each antibody to Ara h 2 and Ara h 3, with resulting affinities of tens of picomolar (pM) to sub-
203 pM for Ara h 2 and tens of nanomolar (nM) to sub-nM for Ara h 3 (Fig. 3C, fig. S8); these
204 affinities are comparable to some of the highest affinity native human antibodies discovered
205 against pathogens such as HIV, influenza, and malaria (41–45). Furthermore, if the antibodies
206 we discovered or variants thereof were to be used therapeutically as blocking antibodies
207 intended to outcompete endogenous IgE for allergen, an approach recently shown to be
208 efficacious for treatment of cat allergy (46), such high affinity to multiple peanut allergens should
209 be advantageous.

210

211 To investigate the degree to which each chain and the mutations therein affect antibody binding
212 properties, we recombinantly expressed eight variants of antibody PA13P1H08, each with one
213 or more regions in the heavy and/or light chain reverted to the inferred naïve rearrangement (fig.
214 S7E-G). Retaining the native heavy chain while swapping the light chain with another kappa
215 light chain from an antibody without peanut allergen specificity abrogated binding to both
216 allergenic proteins, while reverting both chains largely eliminated Ara h 3 binding and
217 dramatically reduced Ara h 2 affinity (Fig. 3C, fig. S8C). Reverting only the heavy or light chain
218 reduced the affinity to Ara h 2 and Ara h 3, but disproportionately; light chain mutations
219 contributed more to Ara h 3 affinity than did heavy chain mutations. We also found a synergistic

220 contribution of heavy chain mutations to affinity as independent reversion of the CDR1, CDR2,
221 or framework regions each caused minor decreases in affinity. Reversion of the CDR3 did not
222 alter binding, which was not surprising given the two amino acid difference between the native
223 and reverted CDR3 sequences. Interestingly, reversion of the heavy chain CDR2 increased Ara
224 h 3 affinity, while only marginally decreasing Ara h 2 affinity. Together, these results indicate
225 that while the inferred naïve antibody is capable of binding the most clinically relevant peanut
226 allergen Ara h 2, mutations in both heavy and light chains are necessary to produce the high
227 affinity and cross-reactive antibodies that we found in circulating IgE PBs of unrelated
228 individuals.

229
230 We also expressed antibodies from two other CFs. CF2 contained three IgE PBs from individual
231 PA16 (two of which were identical), but these antibodies did not bind Ara h 1, 2, or 3, which was
232 unsurprising given this individual had low plasma peanut-specific IgE levels as well as IgE
233 specific to other allergens (fig. S1). On the other hand, CF3 contained an IgE PB (PA15P1D05)
234 and IgG4 PB (PA15P1D12), the recombinantly expressed antibodies from which did not bind
235 Ara h 1 appreciably, but bound Ara h 3 with nanomolar affinity and Ara h 2 with sub-nanomolar
236 affinity (fig. S8). Interestingly, these two antibodies utilize the same light chain V gene and a
237 highly similar heavy chain V gene (IGHV3-30-3*01) as the six convergent antibodies of CF1,
238 which provides support for the importance of these V genes in Ara h 2 and Ara h 3 binding.
239 Moreover, the presence of peanut-specific IgE and IgG4 PBs in the same CF within an allergic
240 individual provides a unique example of *in vivo* competition for allergen between two antibody
241 isotypes with possible antagonistic effector functions in allergic disease.

242
243 **Biallelic and polarized germline transcription in single cells**

244
245 Tailored responses of the adaptive immune system are possible in part due to the ability of
246 activation-induced cytidine deaminase (AID) to initiate class switch recombination (CSR) in B
247 cells, leading to the production of antibodies with specific effector functions. CSR is preceded by
248 cytokine-induced germline transcription, where nonproductive germline transcripts (GLTs) that
249 contain an I-exon, switch (S) region, and heavy chain constant region exons guide AID to the S
250 region (47). Importantly, GLT processing is necessary for CSR (48, 49) and canonically results
251 in two species: an intronic S region lariat and a mature polyadenylated transcript consisting of
252 the I exon spliced to the constant region exons (50). In our scRNA-seq data, we observe
253 multiple splice isoforms of the latter, where the proximal constant region exon serves as the
254 exclusive splice acceptor for multiple splice donors. IgE had the largest number of distinct GLTs
255 at five (Fig. 4A and fig. S9), which we confirmed through Sanger sequencing (fig. S10); these
256 were expressed in numerous cells of varying isotypes and across all individuals, but at
257 nonuniform frequencies. The I-exon was the most common splice donor site (Fig. 4A, GLT #1)
258 and it is known that I-exons can provide multiple splice donors (51–53), but εGLT splice donors
259 within the switch region were also observed. We also found independent evidence for multiple
260 εGLT splice donors in a previously published scRNA-seq dataset from murine B cells harvested
261 24 h after simulation to class switch (54) (fig. S11).

262

263 We next assessed GLT expression across all isotypes. Most B cells did not express a GLT of a
264 non-self isotype, while, unlike previous reports (55), those that did tended to be polarized
265 towards the expression of a single GLT isotype (Fig. 4B). GLT production varied dramatically
266 both by GLT isotype produced and by the cell's current isotype (Fig. 4C). We observed a high
267 proportion of IgG4 and IgE cells expressing εGLTs, while in contrast, we found almost no IgG4
268 GLT expression within any cells in these allergic individuals. Interestingly, we observed that
269 GLT expression arising from the alternate allele is common, as evidenced by widespread
270 expression of IgM GLTs in class switched B cells, and in some cases, expression of GLTs of
271 isotypes upstream of a cell's own class switched isotype (signal below the diagonal in Fig. 4C).
272 Mirroring the landscape of human class switching (56), we observe the trend for GLT production
273 to be higher for proximal downstream isotypes rather than distant downstream isotypes.
274

275 The study of B lymphocyte transcriptomes at single cell resolution offers other advantages; for
276 example, we discovered multiple instances of biallelic GLT expression within single cells
277 through heavy chain constant region haplotype phasing in individuals who had heterozygous
278 single nucleotide variants within these loci. An example of this process that demonstrates
279 biallelic εGLT expression is shown in Fig. 4D. Not all constant regions within individuals had
280 such enabling variants, but in those that did we observed biallelic expression was relatively
281 common relative to monoallelic expression (Fig. 4E).
282

283 Conclusion

284
285 Using scRNA-seq, we provide the first transcriptomic characterization of circulating human IgE
286 B cells and the antibodies they produce. Our data suggest two mechanisms underlying IgE
287 regulation in humans: a relative deficiency of membrane immunoglobulin expression and an
288 immature IgE PB gene expression program suggestive of weakened activation, proliferation,
289 and survival capacity. These results are largely consistent with extensive studies of mIgE
290 signaling and IgE memory in murine models of allergy (57–61), and provide evidence supporting
291 the use of animal models for this disease. Furthermore, the ability to capture GLT splice variant,
292 polarization, and biallelic expression information within single B cells presents an exciting
293 application of scRNA-seq for future mechanistic studies of GLT production and CSR.
294

295 Insight into convergent evolution of high affinity antibodies in unrelated individuals can guide
296 vaccine design and lead to strategies for population-level passive immunity; it is also a process
297 that has been argued to occur in response to a number of pathogens such as influenza (62),
298 HIV (45), *Streptococcus pneumoniae* (63). Here we found a striking case of convergence where
299 two unrelated individuals produced high affinity, cross-reactive, peanut-specific antibodies
300 comprised of identical gene rearrangements within respective heavy and light chains. A third
301 individual had Ara h 2-specific antibodies that utilized a similar heavy V gene and the same light
302 chain V gene. Although our study was limited by sample size, there is evidence supporting the
303 importance of these genes within the peanut-allergic population more broadly: one independent
304 study of IgE heavy chain sequences from peanut allergic individuals (64) reported IgE heavy
305 chains that utilized identical V and J genes and shared at least 70% CDR3 identity with one or

306 more of the six convergent antibodies in our dataset (fig. S12); another study (9) reported Ara h
307 2 specific IgG and IgM antibodies that utilized similar IGHV3-30 genes.
308

309 Cross-inhibition experiments with purified allergens and plasma IgE have shown that cross-
310 reactivity of IgE antibodies may also be common within peanut allergic individuals (65) and the
311 antibodies we have isolated here offer a clear example of these findings. Furthermore, the fact
312 that these high affinity antibodies were being produced by secretory IgE PBs found in circulation
313 contributes to an understanding of how minute amounts of allergen are capable of eliciting
314 severe allergic reactions. We also expect that either these antibodies or engineered variants of
315 them could be used as therapeutic agents; recent clinical results have shown that engineered
316 allergen-specific IgG antibodies can be administered to humans and provide effective treatment
317 for cat allergies, perhaps by outcompeting the native IgE for antigen (46).
318

319 **Acknowledgements**

320 We would like to acknowledge helpful discussions with Felix Horns. This research was
321 supported by the Simons Foundation (SFLIFE #288992 to SRQ), the Chan Zuckerberg Biohub,
322 and the Sean N Parker Center for Allergy and Asthma Research at Stanford University. DC is
323 supported by an NSF Graduate Research Fellowship and the Kou-I Yeh Stanford Graduate
324 Fellowship.
325

326 **Data availability**

327 Processed data including antibody assembly information and the gene expression count matrix
328 is available in the Supplementary Materials. Raw sequencing data is available from the
329 Sequence Read Archive (SRA) under BioProject accession PRJNA472098.
330

331 **Material and Methods**

332 ***Study subjects***

333 All study subjects were consented and screened through a Stanford IRB approved-protocol.
334 Participants were eligible if they had a peanut allergy confirmed by an oral food challenge and
335 board certified allergist. Peanut allergic individuals with reported reactivity to peanut ranged in
336 age from 7 to 16, and in some cases exhibited sensitivities to other food allergens (fig. S1).
337

338 ***Plasma IgE measurement and B cell isolation***

339 Both plasma and cellular fractions were extracted from up to 45 mL of fresh peripheral blood
340 collected in K₂ EDTA tubes. For plasma extraction, blood was transferred to 15 mL falcon tubes
341 and spun at 1600 g for 10 min. The upper plasma layer was extracted, transferred to 2 mL
342 Eppendorf protein LoBind tubes and spun again at 16000 g to further purify the plasma fraction.
343 The resulting supernatant was moved to fresh tubes before being put on dry ice and later
344 transferred to -80°C. Allergen-specific plasma IgE measurements were performed by CLIA-
345 licensed Johns Hopkins University Dermatology, Allergy, and Clinical Immunology (DACI)
346 Reference Laboratory using the ImmunoCAP system. To purify B cells remaining after plasma
347 extraction, RosetteSep human B cell enrichment cocktail (Stemcell Technologies), a negative
348 selection antibody cocktail, was added after the plasma fraction was replaced with PBS + 2%
349 fetal bovine serum (FBS). After a 20 min incubation, the blood was then diluted two-fold with

350 PBS + 2% FBS before being transferred to Sepmate 50 mL tubes (Stemcell Technologies)
351 containing 15 mL Ficoll-Plaque PLUS (GE Healthcare Life Sciences). An enriched B cell
352 population was achieved after a 10 min, 1200 g spin with the brake on and transferred a fresh
353 tube. Residual red blood cells were then removed using ACK lysis buffer (ThermoFisher) and
354 cells were washed with stain buffer (BD Biosciences). Cells were stained on ice with the
355 following BioLegend antibodies according to the manufacturer's instructions: PE anti-human IgE
356 clone MHE-18, Brilliant Violet 421 anti-human CD19 clone HIB19, APC anti-human IgM clone
357 MHM-88, and Alexa Fluor 488 anti-human IgG clone M1310G05. Cells were washed twice more
358 prior to sorting.

359

360 ***Flow cytometry and single cell sorting***

361 Single cell sorts were performed on a FACSaria II Special Order Research Product (BD
362 Biosciences) with a 5 laser configuration (355, 405, 488, 561, and 640 nm excitation).
363 Fluorophore compensation was performed prior to each sort using OneComp eBeads
364 (ThermoFisher), although minimal compensation was required due to the fluorophore panel and
365 laser configuration. Equivalent laser power settings were used for each sort. Cells were sorted
366 using "single cell" purity mode into chilled 96 well plates (Biorad HSP9641) containing 4 μ L of
367 lysis buffer (66) with ERCC synthetic RNA spike-in mix (ThermoFisher). Plates were spun and
368 put on dry ice immediately before storage at -80°C.

369

370 ***cDNA generation, library preparation, and sequencing***

371 A modified version of the Smart-seq2 protocol (11) was used as previously described (66), but
372 with 25 cycles of PCR amplification due to the low mRNA content of naïve and memory B cells.
373 In total, 1165 cells were sequenced across 5 runs using 2x150 bp Illumina High Output kits on
374 an Illumina NextSeq 500.

375

376 ***Sequencing read alignment, gene expression, and splicing***

377 Sequencing reads were aligned to the genome in order to determine gene expression and
378 identify splice variants. To produce the gene expression counts table, reads were first aligned to
379 the GRCh38 human genome using STAR v2.5.3a (12) run in 2-pass mode. Gene counts were
380 then determined using htseq-count (67) run in intersection-nonempty mode. The GTF
381 annotation file supplied to both STAR and htseq-count was the Ensembl 90 release manually
382 cleaned of erroneous immunoglobulin transcripts e.g. those annotated as either a V gene or
383 constant region but containing both V gene and constant region exons. During STAR genome
384 generation an additional splice junction file was provided that included splicing between all
385 combinations of heavy chain CH1 exons and IGHJ genes to improve read mapping across
386 these junctions. Gene expression was normalized using log2 counts per million after removing
387 counts belonging to ERCCs. Cells with fewer than 950 expressed genes were excluded prior to
388 analysis (fig. S3B), as were putative basophils, identified by high FACS IgE, absent or poor
389 quality antibody assemblies, and expression of histidine decarboxylase (HDC) and Charcot-
390 Leyden crystal protein/Galectin-10 (CLC). Batch effects mostly affecting the naïve / memory B
391 cell subset were noted between sorts by clustering using PCA on the 500 most variable genes;
392 this gene set was enriched in genes known to be affected by sample processing such as FOS,
393 FOSB, JUN, JUNB, JUND, HSPA8 (68). PCA following the exclusion of genes differentially

394 expressed between sort batches (Mann-Whitney test, p-value < 0.01 after Bonferroni correction)
395 yielded well-mixed populations within both the naïve / memory and PB cell clusters not biased
396 by sort batch, individual, or sequencing library (fig. S3G). For differential expression analysis
397 between IgE and non-IgE PBs, genes expressed in at least 10 PBs were analyzed by voom-
398 limma (69) with sort batch and sequencing library were supplied as technical covariates.
399 Constant region genes, such as IGHE and IGHA1, were excluded given these are differentially
400 expressed by design of the comparison being made.

401
402 Analysis of splicing, including GLT expression, relied upon splice junctions called by STAR.
403 Junctions were discarded if they contained fewer than three unique reads and GLT splice
404 donors were only considered if observed in at least three cells. It should be noted the elevated
405 levels of IgG2 GLT production can be explained by splicing of the CH1 IgG2 exon to an
406 upstream lincRNA (ENSG00000253364). Biallelic expression of GLTs was determined based
407 on heterozygous expression of single nucleotide variants discovered within heavy chain
408 constant regions using bcftools (70). For the analysis of immunoglobulin heavy chain constant
409 region exon splicing and coverage, genomic coordinates from the Ensembl gene annotation
410 were used. Read coverage of these exons was generated using the samtools (71) depth
411 command. To illustrate the absence of IgE membrane exon coverage, cells were leniently
412 considered to have “any” membrane exon coverage (Fig. 2C-D) if at least 5% of either
413 membrane exon had at least 5 reads.

414
415 **Antibody heavy and light chain assembly**
416 In addition to alignment, sequencing reads were also independently assembled in order to
417 reconstruct full length heavy and light chain transcripts. BASIC (72) was used as the primary
418 assembler given its intended use for antibody reconstruction, while Bridger (73), a *de novo*
419 whole transcriptome assembler, was used to recover the minority of heavy and/or light chains
420 lacking BASIC assemblies. The heavy chain isotype or light chain type (lambda or kappa) was
421 determined using a BLAST (74) database of heavy and light chain constant regions constructed
422 from IMGT sequences (75). Immunoglobulin variable domain gene segment assignment was
423 performed using IgBLAST (76) v1.8.0 using a database of human germline gene segments from
424 IMGT. IgBLAST output was parsed with Change-O and mutation frequency was called with
425 SHazaM (77). Cells without a single productive heavy and single productive light chain, which
426 were all members of the naïve / memory cell cluster, were excluded, leaving a final total of 973
427 cells. Graph-tool (<https://graph-tool.skewed.de/>) was used to draw clonal families and the
428 workflow engine Snakemake (78) was used to execute all analysis pipelines.

429
430 **Recombinant antibody expression**
431 Recombinant expression of select antibodies enabled characterization of antibody specificity
432 and affinity. All heavy chains were expressed as human IgG1, while light chains were expressed
433 as either lambda or kappa as appropriate. Heavy and light chain sequences were synthesized
434 by Genscript after codon optimization and were transiently transfected in HEK293-6E cells.
435 Antibodies were purified with RoboColumn Eshmuno® A columns (EMD Millipore) and were
436 confirmed under reducing and non-reducing conditions by SDS-PAGE and by western blots with

437 goat anti-human IgG-HRP and goat anti-human kappa-HRP or goat anti-human lambda-HRP as
438 appropriate.

439

440 ***Functional antibody characterization***

441 ELISAs were performed to qualitatively assess peanut allergen binding. Purified natural Ara h 1
442 (NA-AH1-1), Ara h 2 (NA-AH2-1) and Ara h 3 (NA-AH3-1), purchased from Indoor
443 Biotechnologies, were immobilized overnight at 4°C using 50 µL at a concentration of 2 ng / uL.
444 Following 3 washes, wells were blocked with 100 µL of PBST (ThermoFisher) + 2% BSA for 2
445 hours. After two washes, 100 µL of primary antibodies were incubated for 2 hours at a
446 concentration of 2 ng / µL in blocking buffer. Following 4 washes, 100 µL of rabbit anti-human
447 HRP (abcam #ab6759) or rabbit anti-mouse HRP (abcam #ab6728) secondary antibodies were
448 incubated for 2 hours at a dilution of 1/1000 in blocking buffer. After 5 washes, 150 µL of 1-Step
449 ABTS Substrate Solution (ThermoFisher) was added to the wells. Color development was
450 measured at 405 nm on a plate reader after 8 - 20 min and reported OD values are after
451 subtraction of signal from no-antibody wells. Negative controls included immobilized BSA as an
452 antigen, as well as a human isotype control primary antibody (abcam #ab206195). One random
453 IgM / IgK antibody we expressed (PA12P4H03) also did not exhibit any binding. Positive
454 controls consisted of monoclonal mouse antibodies 2C12, 1C4, and 1E8 (Indoor
455 Biotechnologies) specific for Ara h 1, Ara h 2, and Ara h 3, respectively.

456

457 Kinetic characterization of antibody interactions with natural purified allergenic peanut proteins
458 was achieved using biolayer interferometry on a ForteBio Octet 96 using anti-human IgG Fc
459 capture (AHC) biosensors with 1X PBST as the assay buffer. The assay was run with the
460 following protocol: up to 600s baseline, 120-150s antibody load, 120-300s baseline,
461 associations of up to 300s, and variable length dissociations that lasted up to 30 min for high
462 affinity antibody-antigen interactions. Biosensors were regenerated by cycling between buffer
463 and pH 1.5 glycine following each experiment. Antibodies were loaded at a concentration of 10-
464 25 nM, while optimal peanut protein concentrations were determined experimentally (fig. S8).
465 Data were processed using ForteBio software using a 1:1 binding model and global fit after
466 reference sensor (ligand, but no analyte) subtraction.

467 **Figure captions**

468

469 **Fig 1. Characterization of single B cells isolated from fresh peripheral blood of allergic**
470 **individuals.** (A) Study overview. (B) Principal component analysis separates naïve / memory
471 (pink) and plasmablast (PB, blue) B cell subsets identified by expression of established
472 transcription factors and marker genes shown in (C). (D) Number of cells belonging to each
473 subtype in (B) by isotype. (E) Isotype, B cell subtype, patient of origin, and mutational frequency
474 of each cell that belongs to a clonal family (CF) with multiple members. CFs referred to in the
475 text are labeled.

476

477 **Fig 2. Characterization of 89 IgE antibodies and the single B cells that produce them.** (A)
478 Phylogenetic depiction of antibody heavy chains arranged by IGHV gene (background color),
479 patient of origin (node color), and mutation frequency (node size). (B) Differential gene
480 expression between IgE PBs and PBs of other isotypes. Positive log fold change indicates
481 genes enriched in IgE PBs. (C) Heavy chain constant region coverage histograms for naïve /
482 memory B cells (top) and PBs (bottom) for select isotypes. Mean normalized read depth and
483 95% confidence interval are indicated by solid lines and shaded area, respectively, for the
484 number of cells (n) inscribed. Heavy chains are oriented in the 5' to 3' direction and membrane
485 exons are the two most 3' exons of each isotype. (D) Summary of (C), but depicting the fraction
486 of cells of each isotype with any membrane exon coverage for both B cell subsets and all
487 isotypes.

488

489 **Fig 3. High affinity, cross-reactive human IgE antibodies.** (A) Highly similar heavy and light
490 chain CDR3s depict convergent evolution in two unrelated individuals (PA12 & PA13). Residues
491 shaded by identity. (B) Indirect ELISA depicting antibody cross-reactivity to multiple peanut
492 allergens. Commercially available mouse monoclonal α Ara h antibodies served as positive
493 controls. (C) Dissociation constants (KDs) to major allergenic peanut proteins Ara h 2 and Ara h
494 3 for each antibody as well as eight variants of PA13P1H08, designated as “heavy – light,”
495 using the abbreviations: N=native, R=reverted, FWRs=framework regions. An “r” prefix indicates
496 reversion of only that region. Binding curves for PA13P1H08 shown above.

497

498 **Fig. 4. Germline transcription in single B cells.** (A) Identification of C ϵ germline transcript
499 splice donors along with the number of cells, by isotype, expressing each. (B) Histogram of the
500 number of non-self GLT isotypes expressed in each cell. (C) Germline transcription heatmap
501 indicating the fraction of cells of a given isotype (rows) expressing a given GLT (column). (D)
502 Example from individual PA11 confirming biallelic ϵ GLT expression in a non-IgE B cell after
503 haplotype phasing the IgE constant region using single IgE B cells. (E) Heatmap indicating the
504 fraction GLT-expressing cells for which expression is biallelic, by individual and GLT isotype.
505 Analysis is limited to constant regions within individuals for which haplotype phasing could be
506 performed.

507 **Supplementary figure captions**

508

509 **Fig. S1. Plasma IgE levels.** (A) Allergen-specific and allergen component (hazelnut, peanut)
510 concentrations. (B) Total IgE concentration. (C) Positive correlation between total plasma IgE
511 concentration and the frequency of IgE B cells among CD19+ B cells. Each point is an
512 individual.

513

514 **Fig. S2. FACS gating and analysis.** (A) Gating strategy for sorting single B cells. IgE+ B cells
515 have been overlaid as red dots. (B) Isotype identity within the final IgE gate as determined by
516 heavy chain transcript assembly. ND=not determined. (C) For reference, putative basophils
517 (CD19- IgE+) display higher IgE surface expression than IgE+ B cells.

518

519 **Fig. S3. scRNA-seq data overview and quality control.** (A) Cells were sequenced in 5
520 libraries to a depth of ~1-2 million reads / cell. (B) Genes per cell histogram. Cells expressing
521 fewer than 950 genes were discarded. (C) Rarefaction curve depicting the number of genes
522 detected as a function of sequencing depth for eight randomly selected cells in each B cell
523 subtype. Solid lines and shaded area represent mean and 95% confidence interval for the gene
524 count, respectively. (D) Read mapping distribution for retained cells. Most reads mapped
525 uniquely (Ensembl gene annotation) and multimapped reads largely belonged to RNA18S5
526 repeats on chr21 and unplaced scaffolds. (E) Read mapping across gene bodies showed
527 minimal 3' or 5' bias. (F) V gene assembly length histogram by chain. (G) PCA on the top 500
528 most variable genes before (top) and after (bottom) batch correction.

529

530 **Fig S4. Auxiliary data supporting B cell subtype classification.** PBs (blue) have greater
531 FACS forward and side scatter (A), more cDNA after Smart-seq2 preamplification (B), and have
532 greater gene expression of antibody light and heavy chain constant regions (C) as compared to
533 the naïve / memory B cell subset (pink). (D) Top differentially expressed genes for each subset.

534

535 **Fig. S5. Additional single cell characterization and analysis of clonal families.** (A) PCA
536 plot as in Fig. 1B, but colored by isotype. (B-D) Analysis of clonal families (CFs). (B) Distribution
537 of the number of cells per CF. (C) Fraction of cells of each isotype that belong to a multi-
538 member CF. (D) Heavy (right, blue) and light (left, red) chain CDR3 sequences and similarity
539 heatmap for CFs in Fig 1E.

540

541 **Fig S6. Antibody comparisons within the IgE isotype and across isotypes.** (A) Heatmap
542 indicating number of IgE antibodies with a given heavy and light chain CDR3 length. (B) Heavy
543 and light chain mutation frequency of each IgE antibody. (C) Silent (S) and replacement (R)
544 mutations by region within IgE heavy and light chains. (D) Heavy chain mutation frequency by
545 isotype. (E) Relative utilization of the lambda and kappa light chain by isotype. (F) Heavy chain
546 V and J gene usage heatmaps by isotype.

547

548 **Fig S7. Sequences, characteristics, and engineered variants of convergent IgE antibodies**
549 **belonging to CF1.** (A-D) Antibody colors are conserved among panels. (A) Heavy chain amino
550 acid sequences and the inferred naïve rearrangement ("Reverted"). Residues shaded by

551 identity. (B) As in (A), but for the light chain. (C) Frequency of silent (S) and replacement (R)
552 mutations by region. (D) Mutation frequency percentiles compared to all class-switched
553 antibodies. (E-G) Engineered variants of PA13P1H08. (E) Native and reverted heavy chain
554 sequences, in addition to sequences where region(s) of the heavy chain have been reverted to
555 the inferred naïve rearrangement. Labels with an “r” prefix indicate only that region has been
556 reverted. FWRs = frameworks. (F) Native and reverted light chain sequences. (G) Sequence of
557 a light chain taken from a random antibody, PA12P4H03, which did not bind any peanut
558 allergens by ELISA.

559

560 **Fig S8. Antibody specificity and affinity measurements.** (A) Ara h 2 binding curves acquired
561 using biolayer interferometry. Each plot depicts a serial two-fold dilution starting with the
562 allergen concentration inscribed in the upper right. (B) As in (A), but for Ara h 3. (C) Indirect
563 ELISA of PA13P1H08 variants. For reference, names are designated as “heavy – light,” using
564 the abbreviations: N=native, R=reverted, FWRs=framework regions. An “r” prefix indicates
565 reversion of only that region.

566

567 **Fig S9. GLT splice donors for all isotypes.** Note that only the first three constant region
568 exons of each isotype are shown for clarity.

569

570 **Fig. S10. Sanger sequencing of five εGLTs amplified from single cell cDNA confirms GLT**
571 **identity and splicing.** Shown for each is of 70 nt of GLT sequence spliced to the first 70 nt of
572 IGHE CH1 in the 5' → 3' orientation. For each GLT, the expected upper sequence agrees with
573 the lower Sanger sequencing result. The same reverse IGHE CH2 primer (5' -
574 TTGATAGTCCCTGGGGTGTACC - 3') was used for all GLTs along with the following forward
575 primers (5' -> 3'): (A) CTGGACTGGGCTGAGCTAGAC, (B-C)
576 GGCCTGAGCTGTGATTGGAAG, (D) CACCCTCACAGCATCAACCAAG, (E)
577 TGCCCGGCACAGAAATAACAAAC.

578

579 **Fig S11. Stimulated murine B cells produce multiple εGLTs.** IGV coverage histograms and
580 splice junctions for the murine ighe constant region in single cells stimulated with IL-4, LPS, and
581 BAFF (54). Arrows indicate unique εGLT splice donors.

582

583 **Fig S12. Similar IgE heavy chain CDR3 sequences in an independent dataset.** Pairwise
584 CDR3 sequence identity of the six convergent heavy chain CDR3 sequences from CF1 of the
585 present study and three IgE heavy chain CDR3 sequences derived from multiple patients in a
586 separate peanut allergy immune repertoire sequencing study (64). Each CDR3 sequence from
587 this separate study shares at least 70% identity with one or more CDR3 sequences from the
588 present study. All sequences share the IGHV3-30 and IGHJ6 gene segments and have CDR3s
589 17 amino acids in length.

590

591 **Supplementary tables**

592

593 **Table S1. Gene expression count matrix.** Each column is a cell and each row is a gene.

594

595 **Table S2. Single cell antibody assemblies.** V, (D), and J gene segment calls as well as
596 isotype and CDR3 amino acid sequence for the heavy and light chains of each cell.
597

598 **References**

- 599 1. C. M. Fitzsimmons, F. H. Falcone, D. W. Dunne, Helminth Allergens, Parasite-Specific
600 IgE, and Its Protective Role in Human Immunity. *Front Immunol.* **5**, 61 (2014).
- 601 2. J. Eckl-Dorna, V. Niederberger, What is the source of serum allergen-specific IgE? *Curr*
602 *Allergy Asthma Rep.* **13**, 281–287 (2013).
- 603 3. W. G. Shreffler, K. Beyer, T.-H. T. Chu, A. W. Burks, H. A. Sampson, Microarray
604 immunoassay: association of clinical history, in vitro IgE function, and heterogeneity of
605 allergenic peanut epitopes. *J Allergy Clin Immunol.* **113**, 776–782 (2004).
- 606 4. D. Croote, S. R. Quake, Food allergen detection by mass spectrometry: the role of
607 systems biology. *npj Syst. Biol. Appl.* **2**, 16022 (2016).
- 608 5. J. J. Heeringa *et al.*, IgE-expressing memory B cells and plasmablasts are increased in
609 blood of children with asthma, food allergy, and atopic dermatitis. *Allergy* (2018),
610 doi:10.1111/all.13421.
- 611 6. M. A. Berkowska *et al.*, Human IgE(+) B cells are derived from T cell-dependent and T
612 cell-independent pathways. *J Allergy Clin Immunol.* **134**, 688–697.e6 (2014).
- 613 7. T. J. Looney *et al.*, Human B-cell isotype switching origins of IgE. *J Allergy Clin Immunol.*
614 **137**, 579–586.e7 (2016).
- 615 8. S. U. Patil *et al.*, Peanut oral immunotherapy transiently expands circulating Ara h 2-
616 specific B cells with a homologous repertoire in unrelated subjects. *J Allergy Clin Immunol.*
617 **136**, 125–134.e12 (2015).
- 618 9. R. A. Hoh *et al.*, Single B-cell deconvolution of peanut-specific antibody responses in
619 allergic patients. *J Allergy Clin Immunol.* **137**, 157–167 (2016).
- 620 10. L. Tordesillas, M. C. Berin, H. A. Sampson, Immunology of food allergy. *Immunity*. **47**, 32–
621 50 (2017).
- 622 11. S. Picelli *et al.*, Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc.* **9**,
623 171–181 (2014).
- 624 12. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21
625 (2013).
- 626 13. S. L. Nutt, P. D. Hodgkin, D. M. Tarlinton, L. M. Corcoran, The generation of antibody-
627 secreting plasma cells. *Nat Rev Immunol.* **15**, 160–171 (2015).
- 628 14. H. Xu *et al.*, Regulation of bifurcating B cell trajectories by mutual antagonism between
629 transcription factors IRF4 and IRF8. *Nat Immunol.* **16**, 1274–1281 (2015).
- 630 15. A. Erazo *et al.*, Unique maturation program of the IgE response in vivo. *Immunity*. **26**,
631 191–203 (2007).
- 632 16. A. Horst *et al.*, Detection and characterization of plasma cells in peripheral blood:
633 correlation of IgE+ plasma cell frequency with IgE serum titre. *Clin Exp Immunol.* **130**,
634 370–378 (2002).
- 635 17. G. Georgiou *et al.*, The promise and challenge of high-throughput sequencing of the
636 antibody repertoire. *Nat Biotechnol.* **32**, 158–168 (2014).
- 637 18. R. Manz, M. Löhning, G. Cassese, A. Thiel, A. Radbruch, Survival of long-lived plasma
638 cells is independent of antigen. *Int Immunol.* **10**, 1703–1711 (1998).
- 639 19. F. Medina, C. Segundo, A. Campos-Caro, I. González-García, J. A. Brieva, The
640 heterogeneity shown by human plasma cells from tonsil, blood, and bone marrow reveals

641 graded stages of increasing maturity, but local profiles of adhesion molecule expression.
642 *Blood*. **99**, 2154–2161 (2002).

643 20. K. L. Calame, K.-I. Lin, C. Tunyaplin, Regulatory mechanisms that determine the
644 development and function of plasma cells. *Annu Rev Immunol.* **21**, 205–230 (2003).

645 21. H. J. Gould, B. J. Sutton, IgE in allergy and asthma today. *Nat Rev Immunol.* **8**, 205–217
646 (2008).

647 22. A. M. Cooper *et al.*, Soluble CD23 controls IgE synthesis and homeostasis in human B
648 cells. *J Immunol.* **188**, 3199–3207 (2012).

649 23. G. Weskamp *et al.*, ADAM10 is a principal “sheddase” of the low-affinity immunoglobulin E
650 receptor CD23. *Nat Immunol.* **7**, 1293–1298 (2006).

651 24. R. Ouchida, T. Kurosaki, J.-Y. Wang, A role for lysosomal-associated protein
652 transmembrane 5 in the negative regulation of surface B cell receptor levels and B cell
653 activation. *J Immunol.* **185**, 294–301 (2010).

654 25. D. Dijkstra, A. Meyer-Bahlburg, Human basophils modulate plasma cell differentiation and
655 maturation. *J Immunol.* **198**, 229–238 (2017).

656 26. J. Hasbold, L. M. Corcoran, D. M. Tarlinton, S. G. Tangye, P. D. Hodgkin, Evidence from
657 the generation of immunoglobulin G-secreting cells that stochastic mechanisms regulate
658 lymphocyte differentiation. *Nat Immunol.* **5**, 55–63 (2004).

659 27. A. Anginot, M. Espeli, L. Chasson, S. J. C. Mancini, C. Schiff, Galectin 1 modulates
660 plasma cell homeostasis and regulates the humoral immune response. *J Immunol.* **190**,
661 5526–5533 (2013).

662 28. W. Leśniak, Ł. P. Słomnicki, A. Filipek, S100A6 - new facts and features. *Biochem
663 Biophys Res Commun.* **390**, 1087–1092 (2009).

664 29. K. Boye, G. M. Maelandsmo, S100A4 and metastasis: a small actor playing many roles.
665 *Am J Pathol.* **176**, 528–535 (2010).

666 30. R. L. Geahlen, Syk and pTyr'd: Signaling through the B cell antigen receptor. *Biochim
667 Biophys Acta.* **1793**, 1115–1127 (2009).

668 31. J. A. Ackermann *et al.*, Syk tyrosine kinase is critical for B cell antibody responses and
669 memory B cell survival. *J Immunol.* **194**, 4650–4656 (2015).

670 32. Z. Yang, B. M. Sullivan, C. D. C. Allen, Fluorescent in vivo detection reveals that IgE(+) B
671 cells are restrained by an intrinsic cell fate predisposition. *Immunity.* **36**, 857–872 (2012).

672 33. M. Poggianella, M. Bestagno, O. R. Burrone, The extracellular membrane-proximal
673 domain of human membrane IgE controls apoptotic signaling of the B cell receptor in the
674 mature B cell line A20. *J Immunol.* **177**, 3597–3605 (2006).

675 34. F. D. Batista, S. Anand, G. Presani, D. G. Efremov, O. R. Burrone, The two membrane
676 isoforms of human IgE assemble into functionally distinct B cell antigen receptors. *J Exp
677 Med.* **184**, 2197–2205 (1996).

678 35. A. Karnowski, G. Achatz-Straussberger, C. Klockenbusch, G. Achatz, M. C. Lamers,
679 Inefficient processing of mRNA for the membrane form of IgE is a genetic mechanism to
680 limit recruitment of IgE-secreting cells. *Eur J Immunol.* **36**, 1917–1925 (2006).

681 36. L. C. Wu, A. A. Zarrin, The production and regulation of IgE by the immune system. *Nat
682 Rev Immunol.* **14**, 247–259 (2014).

683 37. G. Achatz, L. Nitschke, M. C. Lamers, Effect of transmembrane and cytoplasmic domains
684 of IgE on the IgE response. *Science.* **276**, 409–411 (1997).

685 38. S. J. Koppelman, M. Wensing, M. Ertmann, A. C. Knulst, E. F. Knol, Relevance of Ara h1,
686 Ara h2 and Ara h3 in peanut-allergic patients, as determined by immunoglobulin E
687 Western blotting, basophil-histamine release and intracutaneous testing: Ara h2 is the
688 most important peanut allergen. *Clin Exp Allergy*. **34**, 583–590 (2004).

689 39. N. Nicolaou *et al.*, Quantification of specific IgE to whole peanut extract and peanut
690 components in prediction of peanut allergy. *Journal of Allergy and Clinical Immunology*.
691 **127**, 684–685 (2011).

692 40. T. D. Dang *et al.*, Increasing the accuracy of peanut allergy diagnosis by using Ara h 2. *J*
693 *Allergy Clin Immunol*. **129**, 1056–1063 (2012).

694 41. J. Wrammert *et al.*, Rapid cloning of high-affinity human monoclonal antibodies against
695 influenza virus. *Nature*. **453**, 667–671 (2008).

696 42. R. Murugan *et al.*, Clonal selection drives protective memory B cell responses in
697 controlled human malaria infection. *Sci. Immunol.* **3** (2018),
698 doi:10.1126/sciimmunol.aap8029.

699 43. K. Kaur *et al.*, High Affinity Antibodies against Influenza Characterize the Plasmablast
700 Response in SLE Patients After Vaccination. *PLoS ONE*. **10**, e0125618 (2015).

701 44. H. Mouquet *et al.*, Memory B cell antibodies to HIV-1 gp140 cloned from individuals
702 infected with clade A and B viruses. *PLoS ONE*. **6**, e24078 (2011).

703 45. J. F. Scheid *et al.*, Sequence and structural convergence of broad and potent HIV
704 antibodies that mimic CD4 binding. *Science*. **333**, 1633–1637 (2011).

705 46. J. M. Orengo *et al.*, Treating cat allergy with monoclonal IgG antibodies that bind allergen
706 and prevent IgE engagement. *Nat Commun*. **9**, 1421 (2018).

707 47. W. T. Yewdell, J. Chaudhuri, A transcriptional serenAID: the role of noncoding RNAs in
708 class switch recombination. *Int Immunol*. **29**, 183–196 (2017).

709 48. K. Hein *et al.*, Processing of switch transcripts is required for targeting of antibody class
710 switch recombination. *J Exp Med*. **188**, 2369–2374 (1998).

711 49. M. Lorenz, S. Jung, A. Radbruch, Switch transcripts in immunoglobulin class switching.
712 *Science*. **267**, 1825–1828 (1995).

713 50. S. Zheng *et al.*, Non-coding RNA Generated following Lariat Debranching Mediates
714 Targeting of AID to DNA. *Cell*. **161**, 762–773 (2015).

715 51. J. T. Collins, W. A. Dunnick, Germline transcripts of the murine immunoglobulin γ 2a gene:
716 structure and induction by IFN- γ . *Int Immunol*. **5**, 885–891 (1993).

717 52. C. Gaff, S. Gerondakis, RNA splicing generates alternate forms of germline
718 immunoglobulin alpha heavy chain transcripts. *Int Immunol*. **2**, 1143–1148 (1990).

719 53. J. F. Gauchat, D. A. Lebman, R. L. Coffman, H. Gascan, J. E. de Vries, Structure and
720 expression of germline epsilon transcripts in human B cells induced by interleukin 4 to
721 switch to IgE production. *J Exp Med*. **172**, 463–473 (1990).

722 54. Y. L. Wu, M. J. T. Stubbington, M. Daly, S. A. Teichmann, C. Rada, Intrinsic transcriptional
723 heterogeneity in B cells controls early class switching to IgE. *J Exp Med*. **214**, 183–196
724 (2017).

725 55. D. J. Fear, N. McCloskey, B. O'Connor, G. Felsenfeld, H. J. Gould, Transcription of Ig
726 germline genes in single human B cells and the role of cytokines in isotype determination.
727 *J Immunol*. **173**, 4529–4538 (2004).

728 56. F. Horns *et al.*, Lineage tracing of human B cells reveals the in vivo landscape of human
729 antibody class switching. *elife*. **5** (2016), doi:10.7554/elife.16578.

730 57. B. Laffleur *et al.*, Self-Restrained B Cells Arise following Membrane IgE Expression. *Cell*
731 *Rep* (2015), doi:10.1016/j.celrep.2015.01.023.

732 58. K. Haniuda, S. Fukao, T. Kodama, H. Hasegawa, D. Kitamura, Autonomous membrane
733 IgE signaling prevents IgE-memory formation. *Nat Immunol.* **17**, 1109–1117 (2016).

734 59. P. Tong *et al.*, IgH isotype-specific B cell receptor expression influences B cell fate. *Proc*
735 *Natl Acad Sci U S A.* **114**, E8411–E8420 (2017).

736 60. Z. Yang *et al.*, Regulation of B cell fate by chronic activity of the IgE B cell receptor. *elife*. **5**
737 (2016), doi:10.7554/elife.21238.

738 61. B. Laffleur, O. Debeaupuis, Z. Dalloul, M. Cogné, B cell intrinsic mechanisms constraining
739 ige memory. *Front Immunol.* **8**, 1277 (2017).

740 62. K. J. L. Jackson *et al.*, Human responses to influenza vaccination show seroconversion
741 signatures and convergent antibody rearrangements. *Cell Host Microbe*. **16**, 105–114
742 (2014).

743 63. J. Zhou, K. R. Lottenbach, S. J. Barenkamp, A. H. Lucas, D. C. Reason, Recurrent
744 variable region gene usage and somatic mutation in the human antibody response to the
745 capsular polysaccharide of *Streptococcus pneumoniae* type 23F. *Infect Immun.* **70**, 4083–
746 4091 (2002).

747 64. K. Kiyotani *et al.*, Characterization of the B-cell receptor repertoires in peanut allergic
748 subjects undergoing oral immunotherapy. *J Hum Genet.* **63**, 239–248 (2018).

749 65. M. Bublin *et al.*, IgE cross-reactivity between the major peanut allergen Ara h 2 and the
750 nonhomologous allergens Ara h 1 and Ara h 3. *J Allergy Clin Immunol.* **132**, 118–124
751 (2013).

752 66. S. Darmanis *et al.*, Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the
753 Migrating Front of Human Glioblastoma. *Cell Rep.* **21**, 1399–1410 (2017).

754 67. S. Anders, P. T. Pyl, W. Huber, HTSeq — a Python framework to work with high-
755 throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).

756 68. S. C. van den Brink *et al.*, Single-cell sequencing reveals dissociation-induced gene
757 expression in tissue subpopulations. *Nat Methods*. **14**, 935–936 (2017).

758 69. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing
759 and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

760 70. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping
761 and population genetical parameter estimation from sequencing data. *Bioinformatics*. **27**,
762 2987–2993 (2011).

763 71. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–
764 2079 (2009).

765 72. S. Canzar, K. E. Neu, Q. Tang, P. C. Wilson, A. A. Khan, BASIC: BCR assembly from
766 single cells. *Bioinformatics*. **33**, 425–427 (2017).

767 73. Z. Chang *et al.*, Bridger: a new framework for de novo transcriptome assembly using
768 RNA-seq data. *Genome Biol.* **16**, 30 (2015).

769 74. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics*. **10**, 421
770 (2009).

771 75. M.-P. Lefranc *et al.*, IMGT, the international ImMunoGeneTics information system. *Nucleic*
772 *Acids Res.* **37**, D1006–12 (2009).

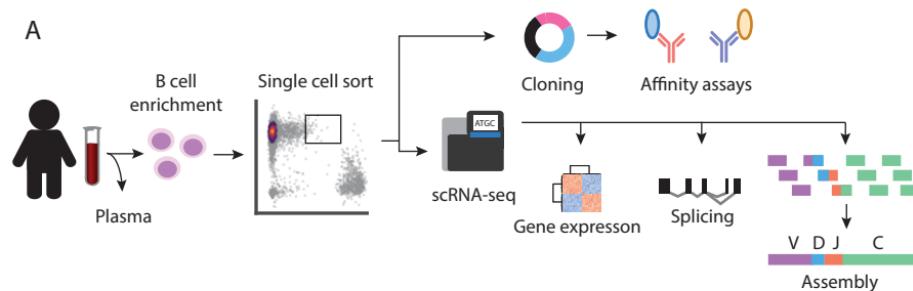
773 76. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: an immunoglobulin variable domain
774 sequence analysis tool. *Nucleic Acids Res.* **41**, W34–40 (2013).

775 77. N. T. Gupta *et al.*, Change-O: a toolkit for analyzing large-scale B cell immunoglobulin
776 repertoire sequencing data. *Bioinformatics*. **31**, 3356–3358 (2015).

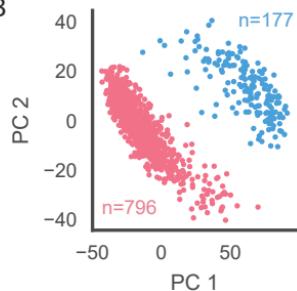
777 78. J. Köster, S. Rahmann, Snakemake--a scalable bioinformatics workflow engine.
778 *Bioinformatics*. **28**, 2520–2522 (2012).

779

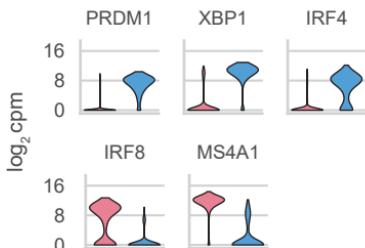
A



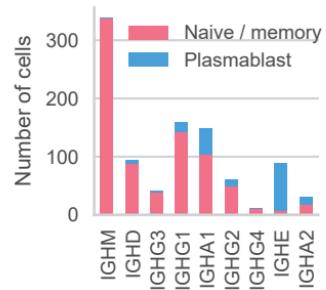
B



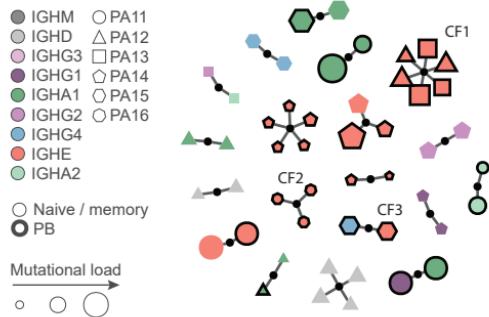
C

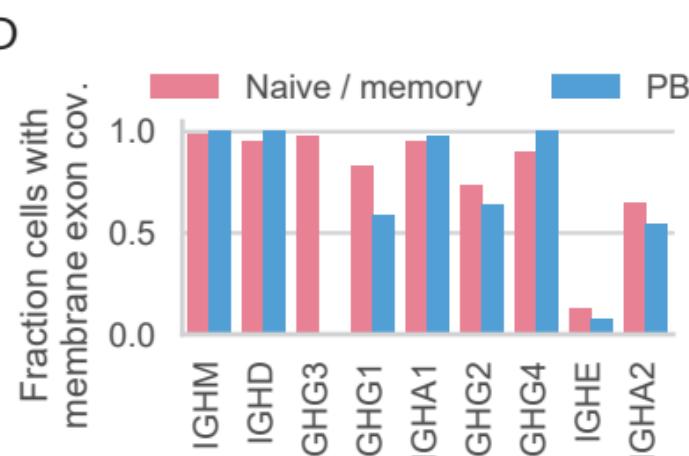
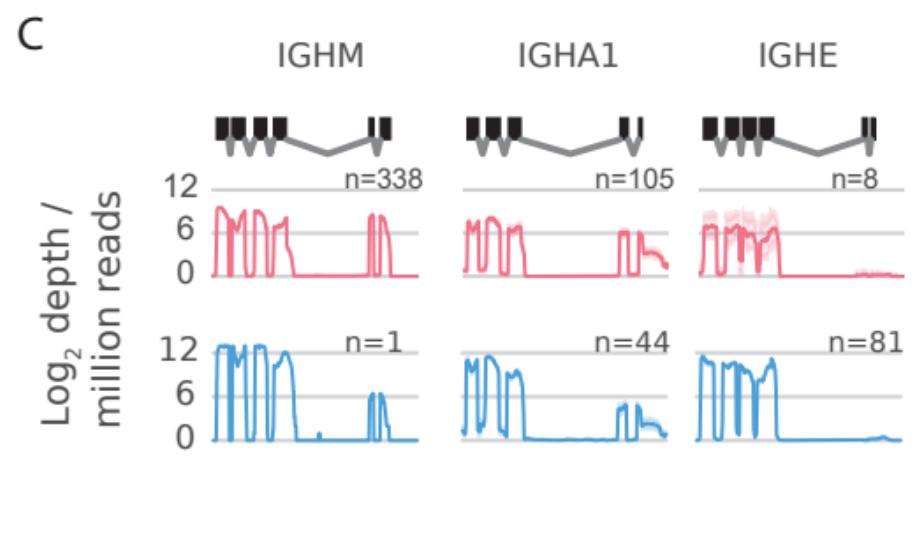
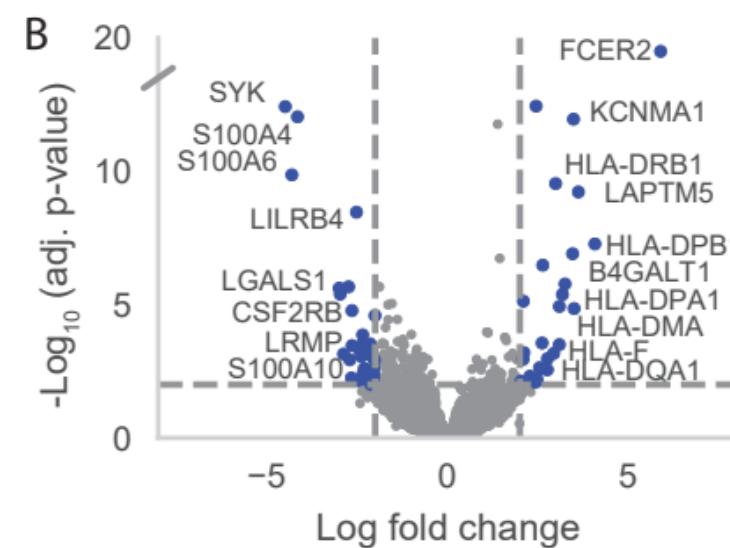
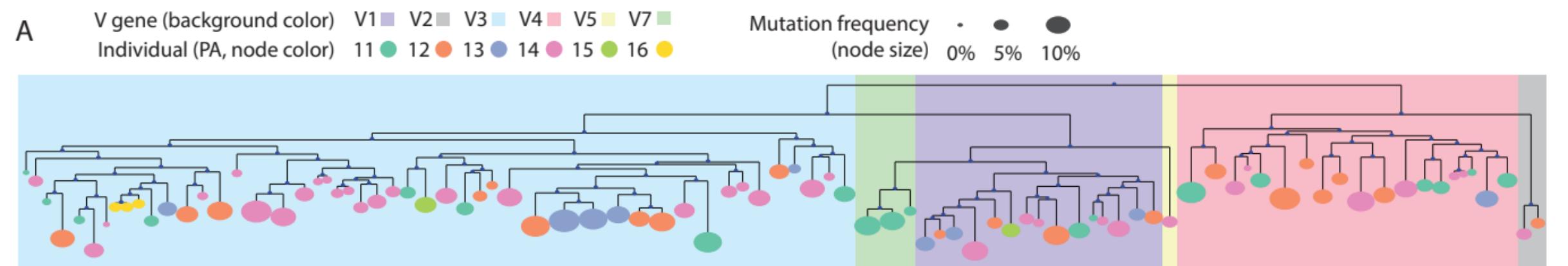


D



E

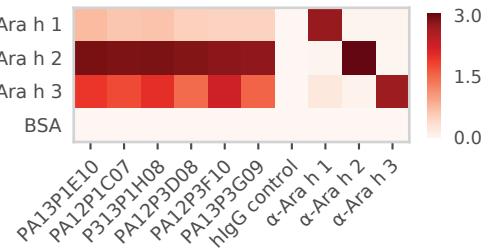




A

CDR3 sequences		
	Heavy chain	Light chain
PA12P3F10	AKVLDYSEFHYYYGLDV	QHYSNSPPYT
PA12P3D08	AKVLDYNEYSLYFGMDV	QYYSDSPPYT
PA12P1C07	AKVLDYSEYSLYFGMDV	QHYSDSPPYT
PA13P3G09	AKVLDYSIFYYYYGLDV	QHYGDSPPYT
PA13P1E10	AKVLDYSAFSYYYGMDV	QHYGRSPPYT
PA13P1H08	AKVLDYSRYSYYYGMDV	QHYGRSPPYT

B



C

