Molecular signatures of non-typeable *Haemophilus influenzae* lung adaptation in paediatric chronic lung disease

Ammar Aziz[1], Derek S. Sarovich[1,2], Elizabeth Nosworthy[1], Jemima Beissbarth[1], Anne B. Chang[1,3], Heidi Smith-Vaughan[1], Erin P. Price[1,2,*,^] and Tegan M. Harris[1,^]

[1]Child Health Division, Menzies School of Health Research, Charles Darwin University, Tiwi, NT, 0810, Australia

[2]GeneCology Research Centre, University of the Sunshine Coast, Sippy Downs, QLD, 4556, Australia

[3]Department of Respiratory and Sleep Medicine, Children's Health Queensland, Queensland University of Technology, Brisbane, QLD, Australia

[^]Authors contributed equally

*Correspondence:

Dr. Erin P. Price

eprice@usc.edu.au

Keywords: Non-typeable *Haemophilus influenzae*, NTHi, RNA-seq, transcriptomics, comparative genomics, convergence, bacterial evolution

Running Title: Signatures of NTHi paediatric lung adaptation

Abstract

Non-typeable *Haemophilus influenzae* (NTHi), an opportunistic pathogen of the upper airways of healthy children, can infect the lower airways, driving chronic lung disease. However, the molecular basis underpinning NTHi transition from a commensal to a pathogen is not clearly understood. Here, we performed comparative genomic and transcriptomic analyses of 12 paired, isogenic NTHi strains, isolated from the nasopharynx (NP) and bronchoalveolar lavage (BAL) of 11 children with chronic lung disease, to identify convergent molecular signatures associated with lung adaptation. Comparative genomic analyses of the 12 NP-BAL pairs demonstrated that five were genetically identical, with the remaining seven differing by only 1 to 3 mutations. Within-patient transcriptomic analyses identified between 2 and 58 differentially expressed genes in 8 of the 12 NP-BAL pairs, including pairs with no observable genomic changes. Whilst no convergence was observed at the gene level; functional enrichment analysis revealed significant under-representation of differentially expressed genes belonging to *Coenzyme metabolism, Function unknown, Translation, ribosomal structure and biogenesis* Cluster of Orthologous Groups categories. In contrast, *Carbohydrate transport and metabolism*, *Cell motility and secretion*, *Intracellular trafficking and secretion*, and *Energy production* categories were over-represented. This observed trend amongst genetically-unrelated NTHi strains provides evidence of convergent transcriptional adaptation of NTHi to paediatric airways that deserves further exploration. Understanding the pathoadaptative mechanisms that NTHi employs to infect and persist in the lower paediatric airways is essential for devising targeted diagnostics and treatments aimed at minimising disease severity, and ultimately, preventing NTHi lung infections and subsequent chronic lung disease in children.

46 **Importance**

47 Non-typeable *Haemophilus influenzae* (NTHi) is a pathogen with increasing public health
48 recognition. Once established in the lower airways, NTHi can facilitate the infection of other
49 pathogenic bacteria, with its presence associated with increased risk of developing
50 bronchiectasis. To elucidate the mechanisms by which NTHi infects the lower airways, we
51 compared the genomes and transcriptomes of 12 paired isolates from 11 children with
52 chronic lung disease. Genomic comparisons identified alterations in several phase-variable
53 and antigenic genes. Unlike NTHi isolated from diseased adult airways, no mutations were
54 observed in *ompP1* in our cohort, suggesting strong stabilising selection in paediatric airways.
55 Transcriptomic analyses identified a parallel signal of pathoadaptation, with over-
56 representation of genes in key functional categories, most of which were down-regulated.
57 These results suggest a complex but ultimately predictable strategy employed by NTHi to
58 colonise paediatric lower airways. Understanding these transcriptional pathoadaptive
59 strategies will enable better treatment and prevention of NTHi lung infections in children.

60 **Introduction**

61 Non-typeable *Haemophilus influenzae* (NTHi) is a Gram-negative bacterium that frequently
62 colonises the respiratory tract mucosa of humans. NTHi is part of the upper airway flora and
63 is found in 20-50% of healthy children and 20-30% of healthy adults (1). The upper airways
64 may act as a reservoir for seeding the lung environment (2), where NTHi switches to an
65 opportunistic pathogenic lifestyle, a transition driven by multiple, poorly understood factors
66 (3).

67 NTHi facilitates the colonisation of other bacterial species in the lower airways (4) and is
68 associated with exacerbations in several respiratory diseases (3). Indeed, its presence in the
69 lower airways is associated with increased risk of future development of bronchiectasis (5).
70 In children with chronic suppurative lung disease (CSLD) or bronchiectasis, NTHi is the most
71 commonly detected bacterium in the lower airways (6). In adults, the bacterium can
72 chronically and repeatedly colonise the lungs of patients with chronic obstructive pulmonary
73 disease (COPD), bronchiectasis, or COPD with bronchiectasis (7, 8), resulting in increased
74 airway inflammation (9). NTHi also has the capacity to form biofilms (10, 11) resulting in a
75 decreased susceptibility to antibiotics and increased inflammatory and defence responses in
76 the host (12). Unsurprisingly, NTHi is a pathogen with increasing public health recognition (4,
77 13).

78 NTHi has evolved several adaptive mechanisms that enable the bacterium to survive and
79 persist in the human host, including modification of its genome via homologous
80 recombination and phase variation (14, 15). Modulation of gene expression through phase
81 variation enables NTHi to rapidly adapt to and colonise new environments (16). Analysis of
82 the NTHi genome has identified several phase-variable virulence factors involved in

83 persistence, nutrient uptake, cellular adherence, and host immune response evasion (17),
84 with direct modulation of gene expression in response to environmental stimuli and host
85 milieu (18). NTHi is naturally competent and can undergo high rates of recombination,
86 resulting in greater genetic and phenotypic diversity than serotypeable *H. influenzae* (19);
87 however, unlike other chronic respiratory infections (20, 21), no significant gene loss or gain
88 has yet been observed in chronic *H. influenzae* infections (17).

89 Identifying signals of NTHi adaptation to the lung would aid in unmasking the mechanisms
90 facilitating pathogenesis, which is critical for improving diagnosis, treatment and prevention
91 of NTHi-associated lower airway diseases. The mechanisms involved in the transition from a
92 commensal state in the nasopharynx (NP) to a pathogenic lifestyle in the lower airways is not
93 well understood (17, 22). In this study, we investigated genome- and transcriptome-wide
94 differences between 12 isogenic NTHi pairs obtained from the NP and lower airways (via
95 bronchoalveolar lavage; BAL) of patients with bronchiectasis or CSLD. We hypothesised that
96 NTHi isolated from the airways of patients with lung disease would exhibit characteristic
97 genetic or transcriptional profiles that signalled a change from a commensal to a pathogen.

98 **Methods**

99 **Isolate Collection.** NP swabs and BAL specimens were collected concurrently from Australian
100 children (mean, 3.2 years; range: 1.1-13.9 years) diagnosed with bronchiectasis or CSLD (Table
101 1) at participating hospitals in the Northern Territory and Queensland (23). Following
102 collection, specimens were transported on ice in 1 mL skim milk, tryptone, glucose and
103 glycerol broth (STGGB) (23, 24) and stored at -80°C before processing. All subsequent cultures
104 were also stored in STGGB at -80°C. Following selective culture on chocolate blood agar (CBA)
105 supplemented with bacitracin, vancomycin, and clindamycin (Oxoid, Scoresby, VIC, Australia),
106 NTHi was isolated and confirmed as previously described (23, 25). Antibiotic sensitivity was
107 determined by disc diffusion as per the Calibrated Dichotomous Susceptibility (CDS) testing
108 guidelines (8th Edition) (26) for ampicillin, amoxicillin-clavulanate, cefaclor, ceftriaxone,
109 chloramphenicol, tetracycline and azithromycin.

110 **Whole-Genome Sequencing and Comparative Genomics.** DNA extraction was performed as
111 previously described (27). Whole-genome sequencing (WGS) was carried out at the Australian
112 Genome Research Facility (Parkville, VIC, Australia) using the Illumina (San Diego, CA, USA)
113 HiSeq 2500 platform. As a criterion for our study, all isolate pairs were required to be highly
114 genetically similar (i.e. isogenic). To identify isogenic NP-BAL pairs from each patient,
115 multilocus sequence typing (MLST) (28) followed by phylogenetic analysis was performed.
116 Further details on isolate selection is described in Text S1. Based on this analysis, we selected
117 12 paired NP-BAL isolates from 11 patients for further investigation; two genetically distinct
118 NP-BAL pairs were selected from one patient (designated 60373_P1 and 60373_P2).

119 To catalogue the mutations separating each NP-BAL pair, SPANDx v3.2 (default settings) (29)
120 was used. Where possible, genetic variants were also confirmed using transcriptomic (RNA-

121 seq) data. For each analysis, the NP isolate draft assembly was used as the reference for
122 variant calling. Further details are provided in Text S1.

123 **NTHi Liquid Media Growth Conditions for RNA Harvest.** Due to the reliance of NTHi growth
124 on the presence of X (haemin/haematin) and V (nicotinamide-adenine-dinucleotide) blood
125 factors (30), we initially attempted to culture isolates in brain heart infusion (BHI) broth
126 supplemented with Haemophilus Test Medium (HTM; Oxoid), and subsequently, a HTM-
127 supplemented artificial sputum medium formulated to mimic the mucosal environment in the
128 airways of cystic fibrosis patients (25, 31). However, due to inconsistent optical density (OD)
129 readings and poor growth or flocculation of some or all isolates in these media, we
130 investigated an alternative liquid medium for the growth of NTHi for RNA-seq (see Text S1 for
131 further details). CBA is widely used for culturing NTHi and other fastidious bacteria (32). Thus,
132 we used this medium as the basis for formulating liquid chocolate medium (LCM).
133 Formulation of LCM and growth conditions are provided in Text S1.

134 **RNA Extraction.** All isolates were cultured in LCM and in duplicate to account for any technical
135 replicability issues. RNA was extracted using TRIzol® as described in Text S1. After DNA digest,
136 absence was verified using real-time PCR assays that detect Hi DNA (33) and defibrinated
137 horse blood DNA (34). For further details see Text S1.

138 **Differential Expression (DE) Analysis.** RNA-seq processing and DE analysis was performed as
139 described in Text S1. Breifly, RNA-seq was carried out using the Illumina HiSeq 2500 platform.
140 In total, 24 isolates were sequenced in duplicate, generating 48 transcriptomes.

141 To identify molecular signatures of lung adaptation on a within-patient basis, DE analysis was
142 first conducted by comparing BAL with NP isolates from individual patients. Subsequently,
143 convergence analysis was performed by comparing BAL with NP isolates from all 12 isolate

144 pairs. The following additional analyses with different sample sets were performed to ensure
145 robustness: a) those NP-BAL pairs with DE detected in the within-patient analysis (eight in
146 total), and b) the three NP-BAL pairs with >45 DE genes (60051, 60068, 60295). For all
147 convergence analyses, the two NP-BAL pairs from patient 60373 were treated either as
148 independent NP-BAL pairs or grouped as a single patient. DE was conducted using edgeR
149 (v3.18.1) (35), implemented in R (v3.4.1) (36). For all analyses, DE was determined using the
150 quasi-likelihood method (*glmQLFit* function).

151 Genes with significant DE were annotated using Clusters of Orthologous Groups (COG) (37),
152 with categories retrieved from a previous study (38) with minor corrections (Table S1). For
153 genes with multiple COG categories, all assigned categories were included in the analysis. To
154 assess for significant enrichment of functional groups, COG categories of the reference
155 genome 86-028NP were compared with those identified in the within-patient DE analyses.
156 Statistical assessment comparing the proportion of categories was performed in R using a
157 two-tailed Fisher's exact test with a false discovery rate corrected *p*-value of ≤0.05.

158 **Data Availability.** All genomic and transcriptomic data generated as part of this study are
159 publicly available under NCBI BioProject PRJNA484075. Genome assemblies for the 24 NTHi
160 strains can be found in NCBI GenBank (SAMN9845260-SAMN9845283), and Illumina WGS and
161 RNA-Seq data are available on the Sequence Read Database (SRR7719256-SRR7719303).

162 **Results**

163 **Comparative Genomic Analysis of Paired Isolates.** To assess diversity and to identify
164 potential clustering of NP or BAL isolates, a maximum parsimony phylogenetic tree was
165 constructed using 124,262 biallelic, core-genome SNPs identified among the paired isolates
166 and a global NTHi isolate set comprising 157 strains (Figure S1). Phylogenomic analysis
167 confirmed that pairs from individual patients clustered closely together yet were distinct from
168 other NP-BAL pairs. Consistent with previous studies (17, 22, 27), there was no evidence of
169 NP- or BAL-specific clades. Isolate pairs were identical sequence types (STs), with 5 of the 12
170 pairs possessing novel STs (ST-1906 to ST-1910; Table S2). Patient 60373 had two distinct
171 isolate pairs (NP-BAL pairs 60373_P1 and 60373_P2) that were separated by >17,000 SNPs,
172 demonstrating the presence of at least two NTHi lineages colonising this patient. This finding
173 was also reflected in the MLST data, with 60373_P1 isolates belonging to ST-1910 whereas
174 60373_P2 isolates were ST-139 (Table S2). Antibiotics are routinely used to treat exacerbation
175 events in paediatric chronic lung disease, and resistance to clinically-relevant antibiotics can
176 arise during such infections. In the current work, no NTHi isolate exhibited resistance towards
177 any of the tested antibiotics, ruling out antibiotic resistance-driven mutations arising between
178 isolate pairs.

179 Next, within-patient comparative genomic analyses were performed to identify variants
180 separating the paired NP-BAL isolates. Pairs were highly genetically similar, with a maximum
181 of three mutations observed between any one pair (Table 1). In total, two non-synonymous
182 SNPs and seven indels (six in open-reading frames) were identified among seven NP-BAL pairs,
183 with the remaining five NP-BAL pairs being genetically identical (Table 2). No copy-number
184 variants, large deletions or inversions were observed between any of the pairs. One non-

185 synonymous SNP occurred in the exogenous haem acquisition-encoding gene, *hgpB*
186 ($HgpB_{Arg407Lys}$; 60295 BAL Hi1), and the second affected the outer membrane protein assembly
187 factor, *bamA* ($BamA_{Glu508Gly}$; 65290 NP Hi3). Of the seven indels, four were simple sequence
188 repeats (SSRs) that were predicted to result in frameshift mutations, leading to truncated
189 proteins in phase-variable genes ($HsdM3_{Glu9fs}$, $HsdM3_{Glu12fs}$, $HgpB_{Thr35fs}$, and $Lic3A_{Ser18fs}$ in
190 60051 BAL Hi1, 60373_P1 BAL Hi1, 60068 BAL Hi1, and 65001 BAL Hi1, respectively). The
191 remaining three indels were in: *ompP2* ($OmpP2_{Ala269_Gly270insValGlyAla}$; 65001 BAL Hi1), *hsdS2*
192 ($HsdS2_{Thr173_Leu176del}$; 65001 BAL Hi1), and an intergenic region 62bp upstream of *hxuC* (60069
193 BAL Hi1). Notably, *hsdM3* (in 60051 and 60373_P1) and *hgpB* (in 60068 and 60295) were
194 mutated in isolates from two patients. When compared with 86-028NP, all mutations were
195 shown to deleteriously affect the BAL isolates (Table 2).

196 The *hsdM3* indels in both the 60051 and 60373_P1 BAL isolates occurred in the same
197 pentanucleotide SSR tract, which comprises variable copy number of a 'CGAGA' motif at the
198 5' end of this gene. Phase variation in *hsdM3* is driven by changes in this highly mutable SSR
199 tract and is influenced by a lack of adenine methylation (39). A 5bp AGACG deletion in this
200 SSR tract in 60051 BAL Hi1 (Figure 1A and 1C) is predicted to produce a truncated protein of
201 just 13 residues. Similarly, a 5bp CGAGA deletion was observed in 60373_P1 BAL Hi1;
202 however, it was only supported by ~60% of aligned reads. Upon closer inspection, a 10bp
203 deletion (CGAGACGAGA) was present in the remaining ~40% of aligned reads (Figure 1B). This
204 mixture of *hsdM3* alleles is supported by the RNA-seq reads. The 5bp (~60%) and 10bp (~40%)
205 components were both predicted to produce a truncated protein of 16 and 24 residues,
206 respectively (Figure 1C). Given that isolates were subjected to a genetic bottleneck prior to
207 sequencing, these mixtures suggest that the *hsdM3* SSR undergoes a rapid mutation rate,
208 with mutations potentially occurring during laboratory passage.

209 **Liquid Chocolate Medium for Reproducible NTHi Growth.** As NTHi has an absolute growth
210 requirement for X and V blood factors (30), commonly used culture media for this fastidious
211 bacterium typically includes lysed blood in a nutrient-rich agar base (e.g. CBA) (32), or BHI
212 base supplemented with HTM (40). However, we found uniform and reproducible growth of
213 NTHi in HTM-supplemented BHI problematic; variable growth rates were observed both
214 among isolates and between technical replicates, indicating inconsistent bacterial growth.
215 Additionally, the microbial load was insufficient for RNA-seq, even after 24h incubation. We
216 also observed inconsistent OD readings between replicates and a tendency towards
217 flocculation in certain strains. Similar issues were observed with HTM-supplemented artificial
218 sputum medium, which is designed to mimic cystic fibrosis sputum (31) (Figure S2A).

219 To overcome these issues, we developed liquid chocolate medium (LCM), a medium that is
220 straightforward to formulate, inexpensive and produces high-quantity NTHi cellular density
221 for downstream analysis. LCM yielded NTHi cultures with reproducible growth, consistent
222 growth rates, little to no flocculation, and sufficient microbial load to support sequencing of
223 late-log phase RNA. LCM was thus used to determine the growth curves of four isolates to
224 identify an appropriate culture duration for RNA extraction. A 7.5-hour post-inoculation time
225 point was found to represent late-log phase growth in all four cultures tested (Figure S2B).
226 This time point, which was chosen to reflect non-logarithmic bacterial growth rates *in vivo*
227 (41, 42), was subsequently used for RNA extraction of all isolates.

228 **Within-patient DE and Comparison with Genetic Mutations.** Culture and RNA extraction of
229 24 isolates was performed in duplicate. On average, 89.2% (87.1-92.0%) of sequence reads
230 aligned uniquely to the 86-028NP reference genome, and of these, an average of 70% (68.2-
231 74.8%) were assigned to known genomic features (Table S2). Due to this high mapping

232 percentage, within-patient DE analysis was performed using 86-028NP as the reference
233 genome. This approach also greatly simplified comparisons among patients. DE was observed
234 between eight NP-BAL pairs, ranging from 2 to 58 DE genes per pair. Three pairs (60051,
235 60068, and 60295) had >45 DE genes, five pairs (60362, 60370, 60373_P1, 60373_P2, and
236 65290) had 15 or fewer DE genes (Dataset S1), and the remaining four pairs (60069, 60294,
237 60316, and 65001) did not exhibit any DE.

238 In total, 189 DE genes, representing a wide functional spectrum, were detected across the
239 eight NP-BAL pairs. Eight genes were DE in three pairs (*comE*, *comM*, *glgC*, *ligA*, *pilB*, *pilC*, *smf*
240 and *tnaA*), 39 were DE in two pairs (*arcC*, *argF*, *asnA*, *comABCDF*, *dmsC*, *glgABPX*, *glpK*, *hitA*,
241 *hsdMS3*, *licBC*, *mao2*, *nagAB*, *nanAEK*, *NTHI0229*, *NTHI0232*, *NTHI0235*, *NTHI0646*, *NTHI0647*,
242 *NTHI1183*, *NTHI1243*, *nrfCD*, *pckA*, *pilA*, *sdaA*, *siaT*, and *tnaB*), and the remaining 87 were DE
243 in single NP-BAL pairs (Figure 2; Dataset S1). However, only ten genes exhibited concordant
244 downregulation (*mao2*, *NTHI0646*, *NTHI0647*, *pckA*, *sdaA*) or upregulation (*asnA, hitA*,
245 *hsdMS3*, *hsdS3, tnaB*) in the BAL isolates. In all other DE genes, discordance in the
246 directionality of expression was observed, including the eight shared DE genes in three NP-
247 BAL pairs (Figure 2). Additionally, genes in the 60295 pair were predominantly upregulated,
248 whereas genes were predominantly downregulated in pairs 60051 and 60068. Overall, there
249 was little overlap amongst patients.

250 There was also minimal correlation between the number of genetic mutations and the
251 number of DE genes detected between NP-BAL pairs (Table 1). For example, the 60295 NP-
252 BAL pair exhibited 46 DE genes, but only a single non-synonymous SNP ($HgpB_{Arg407Lys}$) was
253 observed in the BAL isolate. In contrast, no DE was detected between the 65001 NP-BAL pair
254 (Dataset S1), despite three indels being identified in the BAL isolate, including a deleterious

255 frameshift in the phase-variable sialyltransferase gene, *lic3A*, which truncates the Lic3A
256 protein from 320 to just 36 residues (Table 2). Two NP-BAL pairs (60294 and 60316) contained
257 no mutations and no DE genes. Conversely, two NP-BAL pairs (60051 and 60373_P1)
258 contained both genetic and transcriptional alterations affecting a single phase-variable gene,
259 *hsdM3* (Table 2; Dataset S1). For both pairs, *hsdM3* and *hsdS3* were upregulated in the BAL
260 isolates, with an average $\log_2$ fold change of 2.5 and 4.6, respectively. Additionally, *hsdR3* was
261 upregulated by 2.2-fold in 60373_P1. However, both BAL isolates from 60051 and 60373_P1
262 encode truncated HsdM proteins of 13 and 16-24 residues, respectively (Figure 1C), indicating
263 that *hsdM3* upregulation in these isolates is unlikely to have a functional consequence.
264 Notably, the two isolates with mutations in *hgpB*, 60068 BAL Hi1 and 60295 BAL Hi1, both had
265 a relatively high amount of DE, with 55 and 46 DE genes, respectively.

266 **DE Convergence and Functional Enrichment Analysis of NP-BAL Isolates.** We compared BAL
267 with NP isolates from all patients to identify genes with convergent gene expression.
268 However, no single convergent genetic mutations or DE genes were identified, suggesting
269 more complex pathways involved in lung adaptation. Therefore, we conducted a functional
270 enrichment analysis of the 134 non-redundant DE genes using COG assignments to identify
271 enriched categories. Statistical analysis identified 4/21 COG categories that were significantly
272 over-represented in the functional enrichment analysis (*Carbohydrate transport and*
273 *metabolism*, *Energy production and conversion*, *Intracellular trafficking and secretion*, and
274 *Cell motility and secretion*), and 3/21 categories (*Translation, ribosomal structure and*
275 *biogenesis*, *Coenzyme metabolism*, and *Function unknown*) that were under-represented
276 ($p \leq 0.05$, Figure 3). Of note, the *Function unknown* category encompasses 20.2% genes in the
277 86-028NP genome, yet DE was only observed in 3.5% of these genes. The basis for this under-
278 representation may partly reflect the artificially high number of gene annotations in the NTHi

279 genome. In support of this notion, 47/397 (11.8%) genes assigned to the *Function unknown*
280 category exhibited zero reads across our 48 transcriptomes (Dataset S2, grey shaded rows),
281 representing 37.9% of all genes with zero reads. Approximately 51.8% of the other genes with
282 zero reads corresponded with paralogous ribosomal and transfer RNA species (Dataset S2).

283 The most significant over-represented COGs ($p < 0.001$) were *Carbohydrate transport and*
284 *metabolism*, which is represented by 4.9% of 86-028NP genes, yet 17.3% of DE genes
285 belonged to this category, and *Cell motility and secretion*, with only 0.6% genes in 86-028NP
286 but with 5.8% of DE genes belonging to this COG. The most significant under-represented
287 COGs were *Translation, ribosomal structure and biogenesis*, comprising 8.0% of 86-028NP
288 genes but only 0.6% of the DE genes, and *Function unknown*, which comprises 20.2% of 86-
289 028NP genes yet only 6.2% of DE genes belonged to this category (Figure 3; Dataset S1). Other
290 significant COG categories ($p < 0.01$) were *Energy production and conversion* (5.0% of 86-
291 028NP genes but 12.0% DE genes) and *Coenzyme metabolism* (4.4% 86-028NP genes but 0%
292 DE genes).

293 **Discussion**

294 NTHi is a pathogen of emerging public health importance (4, 43). Conventionally considered
295 a commensal of the upper airways, it is now recognised that NTHi is associated with increased
296 severity and progression of multiple polymicrobial diseases of the lower airways such as
297 COPD, bronchiectasis, and CSLD (44, 45). In this study, we aimed to identify signals of NTHi
298 pathoadaptation to the lower airways using 12 paired isogenic strains retrieved from the
299 upper (NP) and the lower (BAL fluid) airways of 11 paediatric patients with CLSD or
300 bronchiectasis. Similar studies have employed WGS to document the genome-wide changes
301 leading to chronic adaptation and persistence of NTHi retrieved from COPD sputum in adults
302 (17, 46). The current work expands on these prior studies by examining both the genomic and
303 transcriptomic profiles of NTHi adaptation in children's airways. To ensure isolates were
304 retrieved from the lower airways, we obtained NTHi isolates from the lower airways (BAL
305 fluid) rather than sputum, the latter of which is fraught with upper respiratory tract
306 contamination issues (47).

307 As part of our study criteria, only isogenic strains were selected to simplify the identification
308 of convergent pathoadaptative mechanisms. As expected, comparative genomic analyses
309 found minimal genetic differences between all isogenic NP-BAL pairs, with only nine total
310 variants in our dataset, ranging from zero to three variants per pair. Five pairs were genetically
311 identical, with the remaining seven encoding a total of two SNPs and seven indels. Notably,
312 four of these indels occurred in tandem repeat regions, or SSR tracts, of phase-variable genes,
313 which are abundant in the NTHi genome (17, 48). SSRs provide a rapid mechanism for gene
314 expression modulation and adaptation (49, 50). Two SSR indels occurred in the hypervariable
315 Type I restriction-modification (RM) system methyltransferase gene, *hsdM3*. The remaining

316 SSR indels occurred in the haemoglobin-haptoglobin binding protein B-encoding gene *hgpB*,
317 which is involved in NP colonisation and iron acquisition (51), and *lic3A*, which encodes a
318 lipooligosaccharide (LOS) sialyltransferase that facilitates NTHi colonisation in different
319 anatomical sites (52, 53).

320 Our genetic findings concur with the recent studies, which found that phase variation in *lic3A*,
321 *hsdM3* and *hgpB* are common in NTHi isolates retrieved from COPD airways (17, 46). In
322 addition to the *hgpB* SSR variant in 60068 BAL Hi1 that truncates the protein from 992 to 12
323 residues, a non-synonymous SNP ($HgpB_{Arg407Lys}$) was seen in 60295 BAL Hi1. Likewise, the SSR
324 region of *hsdM3* was deleteriously mutated in the 60051 and 60373_P1 BAL isolates, yielding
325 truncated proteins of ≤24 residues (Figure 1C). The *hsdM3* gene is part of a Type I RM system
326 that acts as a defence system against foreign DNA invasion by, e.g. bacteriophages (54), fine-
327 tuning of competence through variable methylation patterns, and in the methylation of self-
328 DNA that may alter gene expression (49). While the *hdsM3* mutations in the paediatric BAL
329 isolates would render their corresponding Type I RM systems non-functional, the phase-
330 variable nature of *hsdM3* enables rapid reverse switching back to an active state. Others have
331 shown that phase-variable RM systems allow NTHi to rapidly generate genomic and
332 phenotypic diversity, which may enhance survival and persistence, especially in new
333 environments (39, 55). Taken together, the commonality of SSR tract mutants affecting *hgpB*,
334 *hsdM3* and *lic3A* in airway-derived isolates demonstrates that phase variation of these loci is
335 driven by strong diversifying selection, with this variation being key in the long-term
336 persistence of NTHi within the hostile lung environment.

337 Outer membrane porin (OmpP1 and OmpP2) mutations are also common in COPD-derived
338 NTHi isolates. Previous studies have identified deleterious *ompP1* mutations in ~33% of COPD

339 isolates, and in-frame *ompP2* mutations (i.e. missense SNPs and indels) in ~20% of COPD
340 isolates (17, 46). Similarly, we identified an in-frame *ompP2* mutation in the 65001 BAL Hi1
341 isolate that occurred in the extracellular loop six motif (56) and increased OmpP2 length by
342 three amino acids. However, we did not observe any genomic or transcriptomic *ompP1*
343 variation between our 12 NP and BAL pairs. Although our sample size is modest, this result
344 suggests that *ompP1* variants may be much less common in paediatric NTHi lung-adapted
345 strains than COPD lung-adapted strains. Biochemical profiles in COPD airways (e.g. fatty acids
346 with bactericidal activity against NTHi wild-type strains) (46) are potentially different to those
347 present in paediatric bronchiectasis or CSLD airways, which may explain the greater selective
348 pressure for OmpP1 inactivation in COPD airways. Further work is needed to document the
349 prevalence of *ompP1* mutations in a larger paediatric lung isolate cohort to examine whether
350 diseased paediatric airways do in fact impart different OmpP1 pressures on NTHi compared
351 with isolates from COPD airways.

352 Across the 12 within-patient DE analyses, a total of 134 non-redundant genes (range: 0 to 58)
353 were DE between NP-BAL pairs, with no correlation between the number of mutations and
354 the number of DE genes. Indeed, NP-BAL pairs 60362, 60370 and 60373_P2 were devoid of
355 mutations, yet DE was observed in 15, 2, and 4 genes, respectively. Likewise, 65001 BAL Hi1
356 encoded mutations in *lic3A*, *ompP2* and a Type I RM system specificity gene, *hsdS2*, yet no DE
357 was detected in this strain. This lack of correlation suggests that DE between isolate pairs is
358 influenced by epigenetic factors (e.g. altered DNA methylation (57)), or by unidentified
359 genetic mutations due to inherent limitations with short-read sequencing approaches (e.g.
360 insertion sequence-mediated rearrangements or variants in paralogues). Future studies
361 employing methylation signature detection and long-read sequencing will enable a deeper
362 understanding of the basis for DE in these strains.

363 We did not identify any DE gene(s) common to all isolates, indicating no universal expression
364 profile associated with lung adaptation (i.e. the BAL isolates). However, some degree of gene
365 convergence was observed, with eight DE genes common to three isolate pairs, and 39
366 common to two pairs (Dataset S1). Of these, the *com* and *pil* operons, required for the
367 biogenesis and function of the type IV pilus, were DE in three pairs. The type IV pilus is
368 involved in multiple biological roles including adherence to epithelial cells, competence,
369 twitching motility, biofilm formation and long-term NP colonisation (58, 59). The pathway
370 responsible for uptake, transport and incorporation of N-Acetylneuraminic acid (Neu5Ac) into
371 LOS (16, 60), was DE in two pairs. Incorporation of Neu5Ac into the LOS enhances immune
372 evasion and increases resistance to killing by human serum (61). Most DE genes (*n*=87)
373 occurred in single pairs, suggesting multiple, parallel strategies towards lower airway
374 adaptation. In addition, the directionality of DE was inconsistent, with gene upregulation in
375 some pairs and downregulation in others (Figure 2). Lack of a convergent transcriptional signal
376 across all strains may be due to sampling difficulties, for example, the inclusion of BAL isolates
377 that have yet to adapt to the lower airways or re-seeding of the NP environment with lung-
378 adapted strains. Greater breadth and depth of strains is needed to determine whether a
379 convergent signal of lung adaptation can be identified in the NTHi transcriptome.

380 To identify parallel signatures of adaptation to the lung environment on a broader scale, we
381 categorised the functions of within-patient DE genes based on COG categories. Overall,
382 strains isolated from the lower airways showed significant enrichment of DE genes in 7/21
383 COG categories. Four categories were over-represented in the DE gene dataset when
384 compared with their prevalence in the 86-028NP genome, with *Carbohydrate transport and*
385 *metabolism*, *Cell motility and secretion*, *Intracellular trafficking and secretion,* and *Energy*
386 *production and conversion* being significant (p <0.05 or p <0.001). In comparison, the COG

387 categories *Translation, ribosomal structure and biogenesis*, *Function unknown* and *Coenzyme*
388 *metabolism* were significantly under-represented (p <0.01 or p <0.001) in the DE gene
389 dataset, suggesting that altered expression of these gene categories is negatively associated
390 with lung adaptation. However, analysis of the 48 transcriptomes revealed that 11.8% genes
391 in the *Function unknown* category had zero reads mapping to the 86-028NP genome (Dataset
392 S2, grey-shaded rows), suggesting that many of these so-called genes likely do not produce
393 messenger RNA or proteins under any growth conditions, and may represent incorrect gene
394 annotations. Further RNA-seq studies of NTHi cultures grown under different conditions will
395 help to resolve these potential annotation issues. Another noteworthy observation was that
396 *Carbohydrate transport and metabolism* and *Energy production and conversion* collectively
397 contained ~35% of all DE genes and were primarily downregulated. This observation is
398 consistent with Baddal and colleagues, who observed the downregulation of genes involved
399 in NTHi central metabolism in response to environmental stimuli (18). *Intracellular trafficking*
400 *and secretion* and *Cell motility and secretion* were also over-represented in the BAL isolates
401 and predominately downregulated (Figure 3; Dataset S1). Taken together, these results
402 suggest that NTHi alters certain functional pathways in response to the lower airway
403 environment, suggesting a degree of predictability in adapting to this new niche. Although
404 further work is needed to consolidate these findings, our work sets the stage for identifying
405 key functional pathways that may be exploited for targeted treatment and eradication of this
406 pathogen.

407 Our study chose not to assess the wider polymicrobial community of the lower airways, as
408 accurate transcriptional characterisation of entire bacterial communities remains
409 bioinformatically challenging (62). We instead chose to focus on an in-depth genomic and
410 transcriptional analysis of NTHi, a species known to play a crucial role in the progression and

411 severity of lung disease (4, 13, 43). Genetically diverse NTHi isolates from different patients
412 were selected to enable identification of convergent adaptation strategies to the lung
413 environment of paediatric patients. Due to the necessary passaging of strains to ensure purity
414 and to obtain sufficient RNA material for sequencing, we cannot discount that these
415 procedures may have influenced the NTHi expression profiles so that they no longer represent
416 their *in vivo* counterparts. However, as previously demonstrated with other bacterial species,
417 isolates appear to maintain their expression profiles even after microbiological handling and
418 culturing *in vitro* (63). For future studies, we recommend maximising the number of strains
419 isolated from individual patients together with longitudinal sampling to better understand
420 pathogen adaptation over time.

421 In conclusion, our study provides new insights into NTHi adaptation to the lower airways in
422 paediatric chronic lung diseases. Through genomic and transcriptomic characterisation of 12
423 paired NTHi isolates from the NP and the lung, we found that NTHi employs several avenues
424 of pathoadaptation that enables this pathogen to persist in the lower airways. Although we
425 did not identify evidence of convergent pathoadaptive mechanisms at the single-gene level,
426 our study identified parallel DE at the functional level, suggesting that NTHi adaptation to the
427 paediatric airways is a complex but ultimately predictable process. The next steps require
428 characterising larger isolate panels using genomic and transcriptomic methods to better
429 understand the pathoadaptive mechanisms that enable NTHi to persist and cause disease.
430 Such findings will be crucial for the informed development of effective therapeutic
431 interventions to prevent NTHi-driven chronic lung diseases.

432 **Conflict of Interest**

433 The authors declare that the research was conducted in the absence of any commercial or
434 financial relationships that could be construed as a potential conflict of interest.

435 **Author Contributions**

436 AC conducted patient recruitment and specimen collection. EN, JB conducted specimen
437 processing, genotyping, and DNA extractions. AA, JB, HSV, EPP, TMH conducted isolate
438 selection. AA conducted culturing, RNA extraction, and bioinformatics analysis with
439 supervision and assistance from DSS, EPP, TMH. AA wrote the initial manuscript draft, and
440 DSS, EPP and TMH critically reviewed and edited the manuscript. DSS, HSV, AC, and EPP
441 conceived of the study and DSS, HSV, and EPP obtained funding. All authors reviewed and
442 approved the final manuscript.

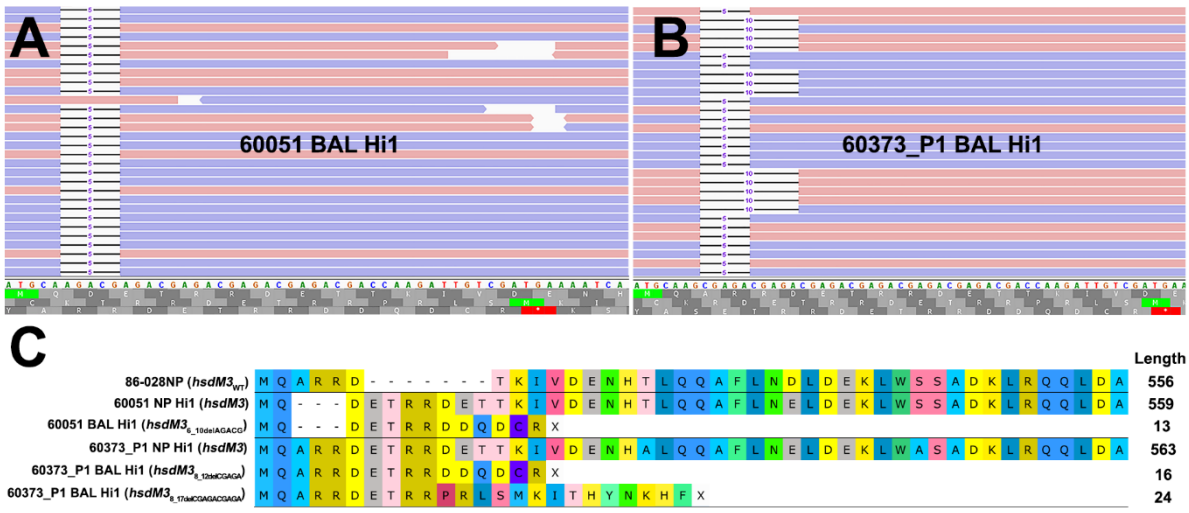
443 **Funding**

444 This work was funded by the Channel 7 Children's Research Foundation (151068) and NHMRC
445 Project Grant (1023781). AA was supported by an RTP scholarship from the Australian
446 Government and an NHMRC CRE scholarship (1078557). DSS was supported by an Advance
447 Queensland fellowship (AQRF13016-17RD2), EPP was supported by a University of the
448 Sunshine Coast fellowship and an Advance Queensland fellowship (AQIRF0362018), and TMH
449 was supported by an NHMRC CRE (1040830). AC is supported by an NHMRC Practitioner
450 Fellowship (APP1154302). HSV was supported by NHMRC CRE Postdoctoral Fellowship
451 (1040830) and NHMRC-funded HOT NORTH collaboration (1131932).

452 **Acknowledgements**

453 We wish to thank the Menzies School of Health Research Child Health team for assistance
454 with microbiological aspects of this study, and Roisin Ure (NHS Greater Glasgow and Clyde,
455 Glasgow, UK) for assistance with *H. influenzae* MLST database submission.

456 **Figures**



457

458 **Figure 1. Read mapping analysis of the highly mutable *hsdM3* pentanucleotide short**
459 **sequence repeat (SSR) tract in 60051 and 60373_P1 NP-BAL pairs.** (A) Graphical
460 representation of 60051 BAL Hi1 Illumina reads aligned against the *hdsM3* region of the 60051
461 NP Hi1 reference assembly. One allele consisting of a 5bp deletion (AGACG) was observed. (B)
462 Graphical representation of 60373_P1 BAL Hi1 Illumina reads aligned against the 60373_P1
463 NP Hi1 reference assembly. Two different alleles consisting of 5bp (AGACG) and 10bp
464 (AGACGx2) deletions were observed in the BAL isolate. (C) Amino acid alignment of *hsdM3*
465 against reference strain 86-028NP (wild-type). NP isolates from 60051 and 60373_P1 encode
466 the full-length protein, with minor differences resulting from the variable SSR track. Indels in
467 the same SSR track of both BAL isolates, including the two alleles observed in 60373_P1, result
468 in a deleterious frameshift that is predicted to yield a truncated protein of 13-24aa.

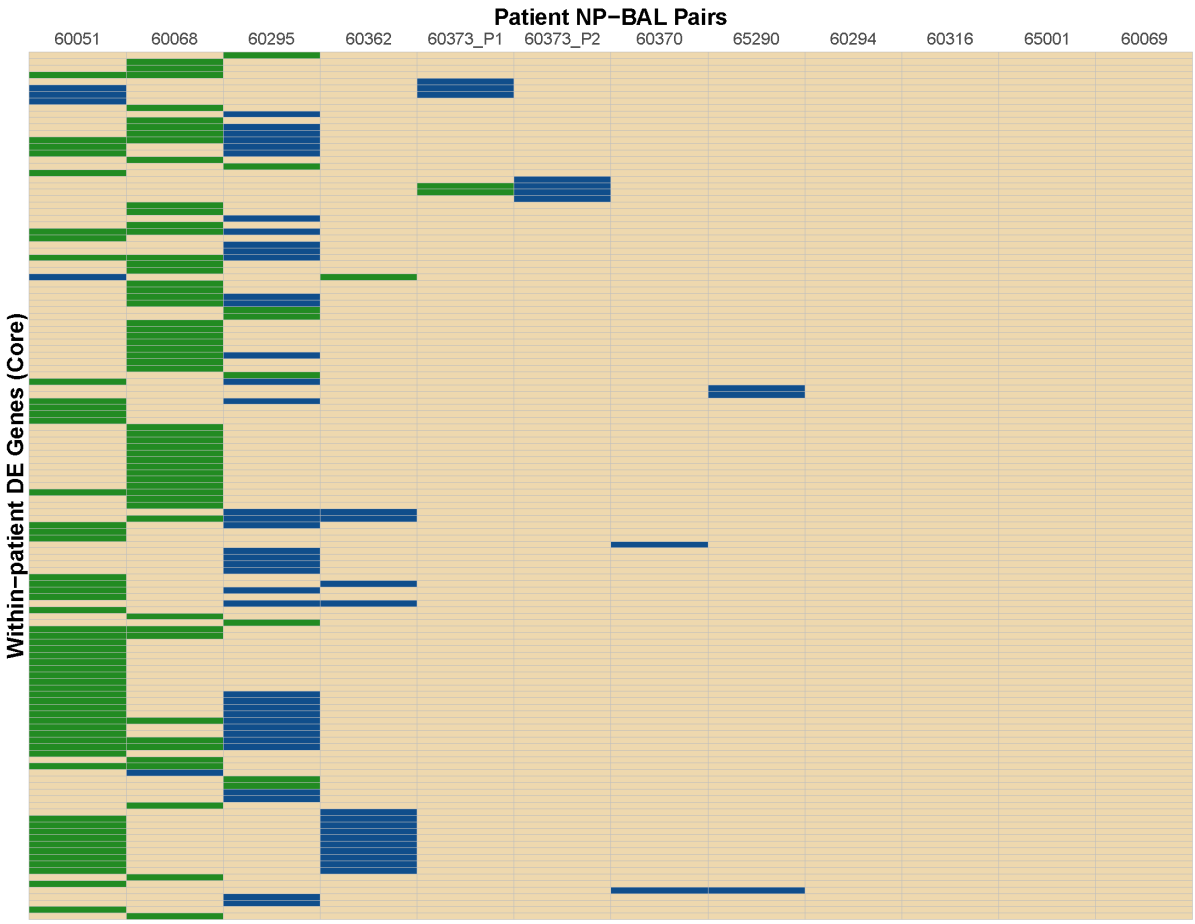


Figure 2. Heat map of differentially expressed (DE) genes identified between the NP-BAL pairs from 11 paediatric chronic lung disease patients. In total 134 non-redundant DE genes were detected across eight NP-BAL pairs. DE was observed in eight NP-BAL pairs; the remaining four pairs contained no DE. Genes were predominantly upregulated in the 60295 NP-BAL pair but predominately downregulated in the 60051 and 60068 NP-BAL pairs. The direction of $\log_2$ fold change of DE genes in BAL is colour-coded as follows: upregulation (blue), downregulation (green), no DE (tan).

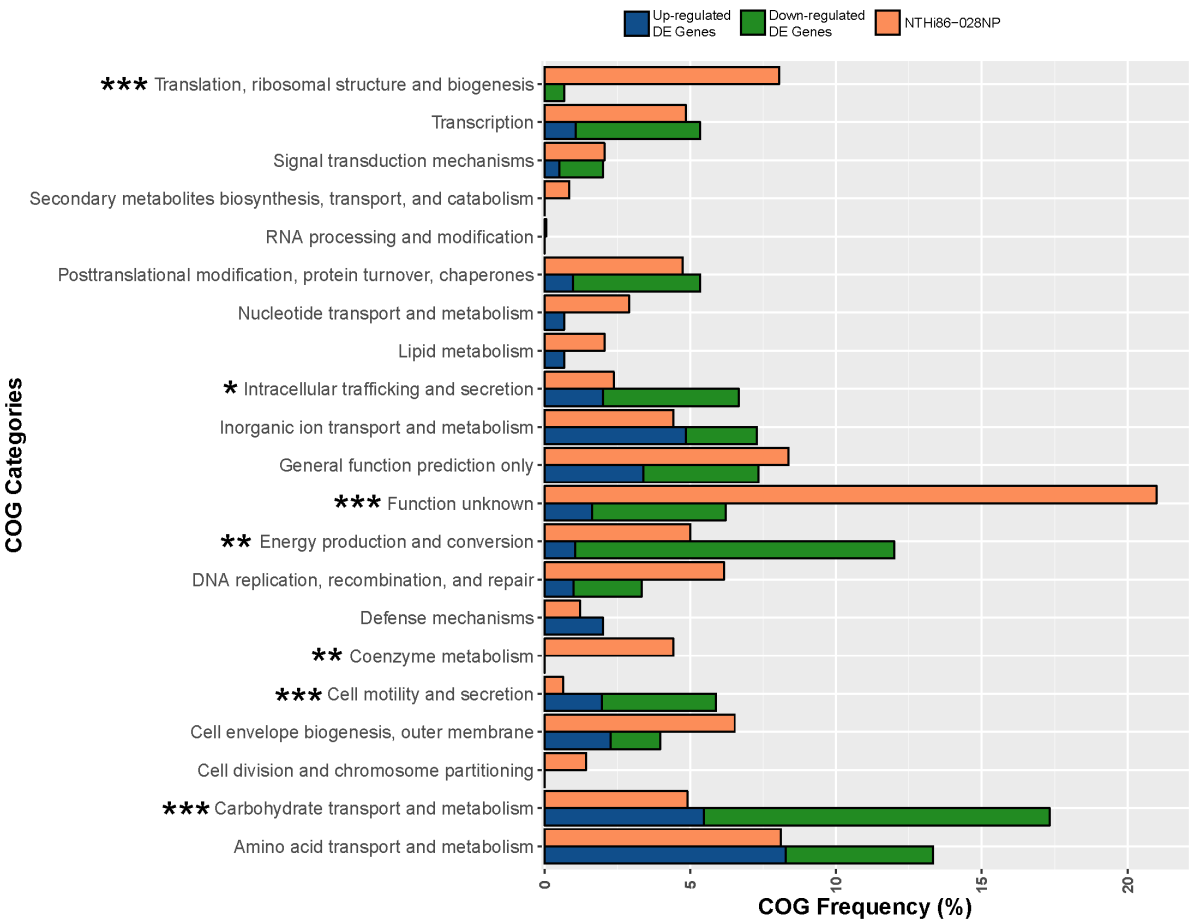


Figure 3. Clusters of Orthologous Group (COG) analysis comparing the frequency of categories in the 86-028NP genome with all DE genes in the 12 NP-BAL pairs. Seven COG categories were significantly enriched in DE genes from BAL-derived isolates, comprising four over-represented categories (*Carbohydrate transport and metabolism, Cell motility and secretion, Energy production and conversion* and *Intracellular trafficking and secretion*) that were majority downregulated and three under-represented categories (*Coenzyme metabolism, Function unknown,* and *Translation, ribosomal structure and biogenesis*). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

486 **Tables**

487 **Table 1.** Summary of genomic and transcriptomic differences between paired, isogenic NTHi
488 isolates obtained from the nasopharynx (NP) and lower airways (bronchoalveolar lavage; BAL)
489 of 11 Australian children with chronic lung disease.

Patient ID	Site of isolation	Isolate no.	No. SNPs	No. indels	No. differentially expressed genes (up/down)	Genes with transcriptional and genetic alterations
60051	NP	1	0	1	58 (4/54)	*hsdM3*
	BAL	1				
60068	NP	3	0	1	54 (1/53)	---
	BAL	1				
60069	NP	2	0	1	0	---
	BAL	1				
60294	NP	1	0	0	0	---
	BAL	1				
60295	NP	1	1	0	46 (38/8)	---
	BAL	1				
60316	NP	1	0	0	0	---
	BAL	1				
60362	NP	1	0	0	15 (14/1)	---
	BAL	2				
60370	NP	2	0	0	2 (2/0)	---
	BAL	2				
60373_P1	NP	1	0	1*	5 (3/2)	*hsdM3*
	BAL	1				
60373_P2	NP	3	0	0	4 (4/0)	---
	BAL	4				
65001	NP	2	0	3	0	---
	BAL	1				
65290	NP	3	1	0	4 (4/0)	---
	BAL	4				

490 Abbreviations: SNP, single-nucleotide polymorphism; indel, insertion-deletion.

491 *The BAL strain from this patient contains a mixture of two deleterious indel mutations in *hsdM3* at a 40:60 ratio (See Figure 1).

492 **Table 2.** Summary of genetic variants identified between 12 paired, isogenic non-typeable *Haemophilus influenzae* isolates retrieved from the
493 nasopharynx vs. lower airways (bronchoalveolar lavage; BAL).

Patient ID[a]	86-028NP Locus Tag	Gene	NP to BAL Genotype		Amino Acid Change (WT Length)	WT Isolate[b]	Mutation Effect in BAL isolate	Affected Protein	Protein Function	COG category
60051	*NTHI1838*	*hsdM3**	AAGACG	A	Glu9fs (556)	NP	Deleterious frameshift; premature stop yields 13aa product	Type I restriction enzyme HindVIIP M protein	DNA binding; site-specific DNA-methyltransferase (adenine-specific) activity	Defense mechanisms
60068	*NTHI0782*	*hgpB**	GCCAACCAA	G	Thr35fs (992)	NP	Deleterious frameshift; premature stop yields 12aa product	Hemoglobin-haptoglobin binding protein B	Receptor and transporter activity	Inorganic ion transport and metabolism
60069	Intergenic	---	GATTATT	G	---	NP	Intergenic (62bp upstream of NTHI0369 [*hxuC*])	---	---	---
60295	*NTHI0782*	*hgpB*	C	T	Arg407Lys (992)	NP	Non-synonymous	Hemoglobin-haptoglobin binding protein B	Receptor and transporter activity	Inorganic ion transport and metabolism
60373_P1†	*NTHI1838*	*hsdM3**	CGAGA (60%); CGAGACGAGA (40%)	A	Glu12fs (556)	NP	Deleterious frameshift; premature stop yields 16aa or 24aa product	Type I restriction enzyme HindVIIP M protein	DNA binding; site-specific DNA-methyltransferase (adenine-specific) activity	Defense mechanisms
65001	*NTHI0472*	*lic3A**	TTTGA	T	Ser18fs (320)	NP	Deleterious frameshift; premature stop yields 36aa product	CMP-Neu5Ac--lipooligosaccharide alpha 2-3 sialyltransferase	Transferase activity; transfer of glycosyl groups	Cell envelope biogenesis
	NTHI0225	*ompP2*	C	CAGCAGTAGG	Ala269_Gly270insValGlyAla (365)	NP	In frame insertion of ValGlyAla; 368aa product	Outer membrane protein P2	Porin activity	Cell envelope biogenesis
	NTHI0315	*hsdS2*	ACTTACCAGCGAG	A	Thr173_Leu176del (390)	NP	In-frame deletion of 4aa	Type I restriction-modification system specificity protein	DNA binding and modification	Defense mechanisms
65290	*NTHI1084*	*bamA*	T	C	Glu508Gly (800)	NP	Non-synonymous	Protective surface antigen D15	Cell outer membrane assembly; protein insertion into membrane	Cell envelope biogenesis

[a]No genetic variants were found between the 60294, 60316, 60362, 60370 and 60373_P2 pairs; only pairs with genetic differences are shown

[b]Wild-type isolate is defined as the isolate which produces full length protein of similar aa length to 86-028NP reference

*Mutation found in a phase-variable region

[†]The BAL strain from this patient contains a mixture of two deleterious indel mutations in *hsdM3* at a 40:60 ratio (See Figure 1).

Abbrevations: aa, amino acid; BAL, bronchoalveolar lavage; COG, cluster of orthologous groups; NP; nasopharynx; WT, wild type

494

495 **Supplemental material legends**

496 **Text S1.** Supplementary methods.

497 **Figure S1.** Maximum parsimony phylogenomic analysis of non-typeable *Haemophilus*
498 *influenzae* isolates from paediatric lung disease against a global strain dataset.

499 **Figure S2.** Non-typeable *Haemophilus influenzae* growth kinetics in different growth media.

500 **Table S1.** Updated Cluster of Orthologous Group terms for non-typeable *Haemophilus*
501 *influenzae* strain 86-028NP. Figshare link: https://doi.org/10.6084/m9.figshare.7956191.v1

502 **Table S2.** Whole-genome sequencing and RNA sequencing summary information. Figshare
503 link: https://doi.org/10.6084/m9.figshare.7956194.v3

504 **Dataset S1.** Within-patient Non-typeable *Haemophilus influenzae* differential expression
505 analysis comparing nasopharynx against bronchoalveolar isolates. Figshare link:
506 https://doi.org/10.6084/m9.figshare.7956158.v1

507 **Dataset S2**. RNA sequencing raw read counts against Non-typeable *Haemophilus influenzae*
508 strain 86-028NP. Figshare link: https://doi.org/10.6084/m9.figshare.7956161.v1

References

1. Slack MPE. 2015. A review of the role of Haemophilus influenzae in community-acquired pneumonia. Pneumonia 6:26-43. doi: 10.15172/pneu.2015.6/520
2. Fothergill JL, Neill DR, Loman N, Winstanley C, Kadioglu A. 2014. *Pseudomonas aeruginosa* adaptation in the nasopharyngeal reservoir leads to migration and persistence in the lungs. Nat Commun 5:4780. doi: 10.1038/ncomms5780
3. Duell BL, Su YC, Riesbeck K. 2016. Host-pathogen interactions of nontypeable *Haemophilus influenzae*: from commensal to pathogen. FEBS Lett 590:3840-3853. doi: 10.1002/1873-3468.12351
4. Van Eldere J, Slack MPE, Ladhani S, Cripps AW. 2014. Non-typeable *Haemophilus influenzae*, an under-recognised pathogen. Lancet Infect Dis 14:1281-1292. doi: 10.1016/s1473-3099(14)70734-0
5. Wurzel DF, Marchant JM, Yerkovich ST, Upham JW, Petsky HL, Smith-Vaughan H, Masters B, Buntain H, Chang AB. 2016. Protracted Bacterial Bronchitis in Children: Natural History and Risk Factors for Bronchiectasis. Chest 150:1101-1108. doi: 10.1016/j.chest.2016.06.030
6. Byun MK, Chang J, Kim HJ, Jeong SH. 2017. Differences of lung microbiome in patients with clinically stable and exacerbated bronchiectasis. PLoS One 12:e0183553. doi: 10.1371/journal.pone.0183553
7. Martinez-Garcia MA, Soler-Cataluna JJ, Donat Sanz Y, Catalan Serra P, Agramunt Lerma M, Ballestin Vicente J, Perpina-Tordera M. 2011. Factors associated with bronchiectasis in patients with COPD. Chest 140:1130-1137. doi: 10.1378/chest.10-1758
8. Sriram KB, Cox AJ, Clancy RL, Slack MPE, Cripps AW. 2018. Nontypeable *Haemophilus influenzae* and chronic obstructive pulmonary disease: a review for clinicians. Crit Rev Microbiol 44:125-142. doi: 10.1080/1040841X.2017.1329274
9. Tufvesson E, Bjermer L, Ekberg M. 2015. Patients with chronic obstructive pulmonary disease and chronically colonized with *Haemophilus influenzae* during stable disease phase have increased airway inflammation. Int J Chronic Obstr 10:881-889. doi: 10.2147/Copd.S78748
10. Murphy TF, Kirkham C. 2002. Biofilm formation by nontypeable *Haemophilus influenzae*: strain variability, outer membrane antigen expression and role of pili. BMC Microbiol 2:7. doi: 10.1186/1471-2180-2-7
11. Swords WE. 2012. Nontypeable *Haemophilus influenzae* biofilms: role in chronic airway infections. Front Cell Infect Microbiol 2:97. doi: 10.3389/fcimb.2012.00097
12. Starner TD, Zhang N, Kim G, Apicella MA, McCray PB, Jr. 2006. *Haemophilus influenzae* forms biofilms on airway epithelia: implications in cystic fibrosis. Am J Respir Crit Care Med 174:213-20. doi: 10.1164/rccm.200509-1459OC
13. Slack MPE. 2017. The evidence for non-typeable *Haemophilus influenzae* as a causative agent of childhood pneumonia. Pneumonia 9. doi: 10.1186/s41479-017-0033-2
14. Power PM, Bentley SD, Parkhill J, Moxon ER, Hood DW. 2012. Investigations into genome diversity of *Haemophilus influenzae* using whole genome sequencing of clinical isolates and laboratory transformants. Bmc Microbiol 12:273. doi: 10.1186/1471-2180-12-273
15. Erwin AL, Nelson KL, Mhlanga-Mutangadura T, Bonthuis PJ, Geelhood JL, Morlin G, Unrath WC, Campos J, Crook DW, Farley MM, Henderson FW, Jacobs RF, Muhlemann K, Satola SW, van Alphen L, Golomb M, Smith AL. 2005. Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. Infect Immun 73:5853-63. doi: 10.1128/IAI.73.9.5853-5863.2005
16. Wong SM, Akerley BJ. 2012. Genome-scale approaches to identify genes essential for *Haemophilus influenzae* pathogenesis. Front Cell Infect Microbiol 2:23. doi: 10.3389/fcimb.2012.00023

17. Pettigrew MM, Ahearn CP, Gent JF, Kong Y, Gallo MC, Munro JB, D'Mello A, Sethi S, Tettelin H, Murphy TF. 2018. *Haemophilus influenzae* genome evolution during persistence in the human airways in chronic obstructive pulmonary disease. Proc Natl Acad Sci U S A doi:10.1073/pnas.1719654115. doi: 10.1073/pnas.1719654115
18. Baddal B, Muzzi A, Censini S, Calogero RA, Torricelli G, Guidotti S, Taddei AR, Covacci A, Pizza M, Rappuoli R, Soriani M, Pezzicoli A. 2015. Dual RNA-seq of Nontypeable *Haemophilus influenzae* and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk. MBio 6:e01765-15. doi: 10.1128/mBio.01765-15
19. Connor TR, Corander J, Hanage WP. 2012. Population subdivision and the detection of recombination in non-typable *Haemophilus influenzae*. Microbiology 158:2958-64. doi: 10.1099/mic.0.063073-0
20. Price EP, Sarovich DS, Mayo M, Tuanyok A, Drees KP, Kaestli M, Beckstrom-Sternberg SM, Babic-Sternberg JS, Kidd TJ, Bell SC, Keim P, Pearson T, Currie BJ. 2013. Within-host evolution of *Burkholderia pseudomallei* over a twelve-year chronic carriage infection. MBio 4(4):e00388-13. doi: 10.1128/mBio.00388-13
21. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, Miller SI, Ramsey BW, Speert DP, Moskowitz SM, Burns JL, Kaul R, Olson MV. 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. Proc Natl Acad Sci U S A 103:8487-92. doi: 10.1073/pnas.0602138103
22. De Chiara M, Hood D, Muzzi A, Pickard DJ, Perkins T, Pizza M, Dougan G, Rappuoli R, Moxon ER, Soriani M, Donati C. 2014. Genome sequencing of disease and carriage isolates of nontypeable *Haemophilus influenzae* identifies discrete population structure. Proc Natl Acad Sci U S A 111:5439-44. doi: 10.1073/pnas.1403353111
23. Hare KM, Grimwood K, Leach AJ, Smith-Vaughan H, Torzillo PJ, Morris PS, Chang AB. 2010. Respiratory bacterial pathogens in the nasopharynx and lower airways of Australian indigenous children with bronchiectasis. J Pediatr 157:1001-5. doi: 10.1016/j.jpeds.2010.06.002
24. Gibson LF, Khoury JT. 1986. Storage and Survival of Bacteria by Ultra-Freeze. Lett Appl Microbiol 3:127-129. doi: 10.1111/j.1472-765X.1986.tb01565.x
25. Price EP, Harris TM, Spargo J, Nosworthy E, Beissbarth J, Chang AB, Smith-Vaughan HC, Sarovich DS. 2017. Simultaneous identification of *Haemophilus influenzae* and *Haemophilus haemolyticus* using real-time PCR. Future Microbiol 12:585-593. doi: 10.2217/fmb-2016-0215
26. Bell SM, Pham JN, Rafferty DL, Allerton JK. 2016. Antibiotic Susceptibility Testing By the CDS Method (8th Edition). http://cdstest.net/wordpress/wp-content/uploads/Antibiotic-Susceptibility-Testing-by-the-CDS-Method-8th-Edition.pdf. Accessed 25/03/2019.
27. Price EP, Sarovich DS, Nosworthy E, Beissbarth J, Marsh RL, Pickering J, Kirkham LA, Keil AD, Chang AB, Smith-Vaughan HC. 2015. *Haemophilus influenzae*: using comparative genomics to accurately identify a highly recombinogenic human pathogen. BMC Genomics 16:641. doi: 10.1186/s12864-015-1857-x
28. Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. doi: 10.1186/1471-2105-11-595
29. Sarovich DS, Price EP. 2014. SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. BMC Res Notes 7:618. doi: 10.1186/1756-0500-7-618
30. Holt LB. 1962. The growth-factor requirements of *Haemopiius influenzae*. J Gen Microbiol 27:317-22. doi: 10.1099/00221287-27-2-317
31. Fung C, Naughton S, Turnbull L, Tingpej P, Rose B, Arthur J, Hu H, Harmer C, Harbour C, Hassett DJ, Whitchurch CB, Manos J. 2010. Gene expression of *Pseudomonas aeruginosa* in a mucin-containing synthetic growth medium mimicking cystic fibrosis lung sputum. J Med Microbiol 59:1089-100. doi: 10.1099/jmm.0.019984-0

32. Artman M, Domenech E, Weiner M. 1983. Growth of *Haemophilus influenzae* in simulated blood cultures supplemented with hemin and NAD. J Clin Microbiol 18:376-9. doi:
33. Aziz A, Sarovich DS, Harris TM, Kaestli M, McRobb E, Mayo M, Currie BJ, Price EP. 2017. Suspected cases of intracontinental *Burkholderia pseudomallei* sequence type homoplasy resolved using whole-genome sequencing. Microb Genom 3. doi: 10.1099/mgen.0.000139
34. Humair PF, Douet V, Moran Cadenas F, Schouls LM, Van De Pol I, Gern L. 2007. Molecular identification of bloodmeal source in *Ixodes ricinus* ticks using 12S rDNA as a genetic marker. J Med Entomol 44:869-80. doi: 10.1093/jmedent/44.5.869
35. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-40. doi: 10.1093/bioinformatics/btp616
36. R Core Team. 2014. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
37. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261-9. doi: 10.1093/nar/gku1223
38. Santana EA, Harrison A, Zhang XJ, Baker BD, Kelly BJ, White P, Liu YL, Munson RS. 2014. HrrF Is the Fur-Regulated Small RNA in Nontypeable *Haemophilus influenzae*. Plos One 9. doi: 10.1371/journal.pone.0105644
39. Zaleski P, Wojciechowski M, Piekarowicz A. 2005. The role of Dam methylation in phase variation of *Haemophilus influenzae* genes involved in defence against phage infection. Microbiology 151:3361-9. doi: 10.1099/mic.0.28184-0
40. Coleman HN, Daines DA, Jarisch J, Smith AL. 2003. Chemically defined media for growth of *Haemophilus influenzae* strains. J Clin Microbiol 41:4408-10. doi: 10.1128/JCM.41.9.4408-4410.2003
41. Kasi AS, Neubauer C, Kato RM, Newman DK. 2017. Bacterial Growth Rate In Cystic Fibrosis Pulmonary Exacerbation. Am J Respir Crit Care Med 195. doi: 10.1164/ajrccm-conference.2017.195.1_MeetingAbstracts.A4856
42. Yang L, Haagensen JA, Jelsbak L, Johansen HK, Sternberg C, Hoiby N, Molin S. 2008. *In situ* growth rates and biofilm development of *Pseudomonas aeruginosa* populations in chronic lung infections. J Bacteriol 190:2767-76. doi: 10.1128/JB.01581-07
43. Erwin AL, Smith AL. 2007. Nontypeable *Haemophilus influenzae*: understanding virulence and commensal behavior. Trends Microbiol 15:355-62. doi: 10.1016/j.tim.2007.06.004
44. Murphy TF. 2003. Respiratory infections caused by non-typeable *Haemophilus influenzae*. Curr Opin Infect Dis 16:129-34. doi: 10.1097/01.aco.0000065079.06965.e0
45. Purcell P, Jary H, Perry A, Perry JD, Stewart CJ, Nelson A, Lanyon C, Smith DL, Cummings SP, De Soyza A. 2014. Polymicrobial airway bacterial communities in adult bronchiectasis patients. BMC Microbiol 14:130. doi: 10.1186/1471-2180-14-130
46. Moleres J, Fernandez-Calvet A, Ehrlich RL, Marti S, Perez-Regidor L, Euba B, Rodriguez-Arce I, Balashov S, Cuevas E, Linares J, Ardanuy C, Martin-Santamaria S, Ehrlich GD, Mell JC, Garmendia J. 2018. Antagonistic Pleiotropy in the Bifunctional Surface Protein FadL (*OmpP1*) during Adaptation of *Haemophilus influenzae* to Chronic Lung Infection Associated with Chronic Obstructive Pulmonary Disease. MBio 9. doi: 10.1128/mBio.01176-18
47. Grønseth R, Drengenes C, Wiker HG, Tangedal S, Xue Y, Husebø GR, Svanes Ø, Lehmann S, Aardal M, Hoang T, Kalananthan T, Hjellestad Martinsen EM, Orvedal Leiten E, Aanerud M, Nordeide E, Haaland I, Jonassen I, Bakke P, Eagan T. 2017. Protected sampling is preferable in bronchoscopic studies of the airway microbiome. ERJ Open Research 3:00019-2017. doi: 10.1183/23120541.00019-2017
48. Gilsdorf JR, Marrs CF, Foxman B. 2004. *Haemophilus influenzae:* Genetic Variability and Natural Selection To Identify Virulence Factors. Infect Immun 72:2457-2461. doi: 10.1128/iai.72.5.2457-2461.2004

49. Vasu K, Nagaraja V. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev 77:53-72. doi: 10.1128/MMBR.00044-12
50. De Ste Croix M, Vacca I, Kwun MJ, Ralph JD, Bentley SD, Haigh R, Croucher NJ, Oggioni MR. 2017. Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. FEMS Microbiol Rev 41:S3-S15. doi: 10.1093/femsre/fux025
51. Poole J, Foster E, Chaloner K, Hunt J, Jennings MP, Bair T, Knudtson K, Christensen E, Munson RS, Jr., Winokur PL, Apicella MA. 2013. Analysis of nontypeable *Haemophilus influenzae* phase-variable genes during experimental human nasopharyngeal colonization. J Infect Dis 208:720-7. doi: 10.1093/infdis/jit240
52. Hosking SL, Craig JE, High NJ. 1999. Phase variation of *lic1A*, *lic2A* and *lic3A* in colonization of the nasopharynx, bloodstream and cerebrospinal fluid by *Haemophilus influenzae* type b. Microbiology 145(11):3005-11. doi: 10.1099/00221287-145-11-3005
53. Hood DW, Cox AD, Gilbert M, Makepeace K, Walsh S, Deadman ME, Cody A, Martin A, Mansson M, Schweda EK, Brisson JR, Richards JC, Moxon ER, Wakarchuk WW. 2001. Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from *Haemophilus influenzae*. Mol Microbiol 39:341-50. doi: doi.org/10.1046/j.1365-2958.2001.02204.x
54. Murray NE. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev 64:412-34. doi: 10.1128/MMBR.64.2.412-434.2000
55. Atack JM, Srikhanta YN, Fox KL, Jurcisek JA, Brockman KL, Clark TA, Boitano M, Power PM, Jen FEC, McEwan AG, Grimmond SM, Smith AL, Barenkamp SJ, Korlach J, Bakaletz LO, Jennings MP. 2015. A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. Nat Commun 6. doi: 10.1038/ncomms8828
56. Sikkema DJ, Murphy TF. 1992. Molecular analysis of the P2 porin protein of nontypeable *Haemophilus influenzae*. Infect Immun 60:5204-11. doi:
57. Fox KL, Dowideit SJ, Erwin AL, Srikhanta YN, Smith AL, Jennings MP. 2007. *Haemophilus influenzae* phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. Nucleic Acids Research 35:5242-5252. doi: 10.1093/nar/gkm571
58. Bakaletz LO, Baker BD, Jurcisek JA, Harrison A, Novotny LA, Bookwalter JE, Mungur R, Munson RS. 2005. Demonstration of type IV pilus expression and a twitching phenotype by *Haemophilus influenzae*. Infect Immun 73:1635-1643. doi: 10.1128/IAI.73.3.1635-1643.2005
59. Jurcisek JA, Bookwalter JE, Baker BD, Fernandez S, Novotny LA, Munson RS, Jr., Bakaletz LO. 2007. The PilA protein of non-typeable *Haemophilus influenzae* plays a role in biofilm formation, adherence to epithelial cells and colonization of the mammalian upper respiratory tract. Mol Microbiol 65:1288-99. doi: 10.1111/j.1365-2958.2007.05864.x
60. Apicella MA. 2012. Nontypeable *Haemophilus influenzae*: the role of N-acetyl-5-neuraminic acid in biology. Front Cell Infect Microbiol 2:19. doi: 10.3389/fcimb.2012.00019
61. Mandrell RE, McLaughlin R, Aba Kwaik Y, Lesse A, Yamasaki R, Gibson B, Spinola SM, Apicella MA. 1992. Lipooligosaccharides (LOS) of some *Haemophilus* species mimic human glycosphingolipids, and some LOS are sialylated. Infect Immun 60:1322-8. doi:
62. Lim YW, Schmieder R, Haynes M, Willner D, Furlan M, Youle M, Abbott K, Edwards R, Evangelista J, Conrad D, Rohwer F. 2013. Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. J Cyst Fibros 12:154-64. doi: 10.1016/j.jcf.2012.07.009
63. Viberg LT, Sarovich DS, Kidd TJ, Geake JB, Bell SC, Currie BJ, Price EP. 2017. Within-Host Evolution of *Burkholderia pseudomallei* during Chronic Infection of seven Australasian cystic fibrosis patients. MBio 8:e00356-17. doi: 10.1128/mBio.00356-17