

# Whole-genome sequencing analysis of clozapine-induced myocarditis

Running title: Genome analysis of clozapine-induced myocarditis

Ankita Narang<sup>1</sup>, Paul Lacaze<sup>2</sup>, Kathlyn J Ronaldson<sup>2</sup>, John J McNeil<sup>2</sup>, Mahesh Jayaram<sup>3,4</sup>, Naveen Thomas<sup>3,4</sup>, Rory Sellmer<sup>5</sup>, David Crockford<sup>5</sup>, Robert Stowe<sup>6</sup>, Steven C. Greenway<sup>7,8,9,10,11</sup>, Christos Pantelis<sup>3,4</sup>, Chad A Bousman<sup>1,5,9,12,13</sup>

<sup>1</sup>Department of Medical Genetics, University of Calgary, Calgary, AB, Canada.

<sup>2</sup>Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia

<sup>3</sup>Adult Mental Health Rehabilitation Unit, North Western Mental Health, Melbourne Health, Western Health, Sunshine Hospital, St Albans, VIC, Australia

<sup>4</sup>Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, Western Centre for Health & Education, St Albans, VIC, Australia

<sup>5</sup>Department of Psychiatry, University of Calgary, Calgary, AB, Canada

<sup>6</sup>Departments of Psychiatry and Neurology (Medicine), Neuropsychiatry Program, and Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC, Canada

<sup>7</sup>Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>8</sup>Department of Biochemistry & Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>9</sup>Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>10</sup>Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>11</sup>Libin Cardiovascular Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>12</sup>Department of Physiology & Pharmacology, University of Calgary, Calgary, AB, Canada.

<sup>13</sup>Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

**Correspondence:** Chad Bousman, Department of Medical Genetics, University of Calgary, 270 HMRB, 3330 Hospital Drive NW, Calgary, AB T2N 4N1, Canada. Tel: +1 403 210 7273;

e-mail: [chad.bousman@ucalgary.ca](mailto:chad.bousman@ucalgary.ca)

38 **ABSTRACT**

39

40 One of the concerns limiting the use of clozapine in schizophrenia treatment is the risk of rare but  
41 potentially fatal myocarditis. Our previous genome-wide association study and human leucocyte  
42 antigen analyses identified putative loci associated with clozapine-induced myocarditis. However, the  
43 contribution of DNA variation in cytochrome P450 genes, copy number variants and rare deleterious  
44 variants have not been investigated. We explored these unexplored classes of DNA variation using  
45 whole-genome sequencing data from 25 cases with clozapine-induced myocarditis and 25  
46 demographically-matched clozapine-tolerant control subjects. We identified 15 genes based on rare  
47 variant gene-burden analysis (*MLLT6*, *CADPS*, *TACC2*, *L3MBTL4*, *NPY*, *SLC25A21*, *PARVB*,  
48 *GPR179*, *ACAD9*, *NOL8*, *C5orf33*, *FAM127A*, *AFDN*, *SLC6A11*, *PXDN*) nominally associated  
49 ( $p < 0.05$ ) with clozapine-induced myocarditis. Of these genes, 13 were expressed in human myocardial  
50 tissue. Although independent replication of these findings is required, our study provides preliminary  
51 insights into the potential role of rare genetic variants in susceptibility to clozapine-induced  
52 myocarditis.

53

54

55 **INTRODUCTON**

56 Clozapine initiation for patients suffering from treatment-resistant schizophrenia is recommended at  
57 the earliest possible opportunity by clinical guidelines,<sup>1, 2</sup> and there is evidence for its superior efficacy  
58 in this difficult-to-treat population compared with other pharmacological treatments<sup>3</sup>. Regulatory  
59 requirements around serious adverse event monitoring have largely addressed the risk of undiagnosed  
60 agranulocytosis. However, there has been ongoing concern about a serious idiosyncratic adverse drug  
61 reaction (ADR) involving inflammation of the myocardium.

62 In 1999, Kilian and colleagues<sup>4</sup> reported the rate of myocarditis (both fatal and non-fatal) as 1.9  
63 per 1000 patients starting clozapine. However, more recent projections suggest rates as high as 3% in  
64 the context of systematic monitoring, with fatality rates ranging from 7% to 75%<sup>5</sup>. Despite these  
65 alarming rates, our current ability to identify those at greatest risk for clozapine-induced myocarditis is  
66 inadequate. This uncertainty has led to reduced clozapine prescription and arguably poorer patient  
67 outcomes. Although research to uncover risk factors has identified a few clinical markers<sup>6,7</sup>, these  
68 factors do not account for all the risk. As such, identification of additional markers is needed to reduce  
69 or prevent fatalities and morbidity resulting from clozapine-induced myocarditis, potentially  
70 facilitating expanded and safer use of this highly effective medication.

71 Genomic analysis is a promising and largely unexplored approach for identifying those at risk  
72 of idiosyncratic ADRs. In fact, our recent genome-wide association study (GWAS) and human  
73 leucocyte antigen (HLA) analyses of clozapine-induced myocarditis identified four independent loci  
74 (*SIX3*, *EFHC1*, *ADAM7*, and *GNAI5*) suggestive of increased myocarditis risk ( $P < 1 \times 10^{-6}$ ), with  
75 odds ratios ranging from 5.5 to 13.7. We also found that the HLA-C\*07:01 allele was associated with  
76 2.89-fold (95% CI: 1.11–7.53) greater odds of clozapine-induced myocarditis<sup>8</sup>. To build on these  
77 promising findings, we undertook the first whole-genome sequencing study to examine different  
78 classes of DNA variation not measured in our previous genetic study, including rare deleterious coding  
79 variants, copy number variants, and candidate variation in pharmacogenes associated with clozapine  
80 metabolism. Herein, we report our results among a cohort of individuals with and without a history of  
81 clozapine-induced myocarditis, in collaboration with members of the Pharmacogenomics of  
82 Clozapine-Induced Myocarditis (PROCLAIM) Consortium<sup>9</sup>.

83

84

85

86

## 87 **METHODS**

### 88 **Study Participants**

89 Participants with (n = 25) and without (n = 25) a history of clozapine-induced myocarditis were  
90 recruited by investigators affiliated with the PROCLAIM Consortium at the University of Calgary,  
91 University of Melbourne, and Monash University. For this study, we included 34 participants (19 cases  
92 and 19 controls) included in our previous GWAS study<sup>8</sup>. For each recruited case of clozapine-induced  
93 myocarditis, an age ( $\pm$  5 years), sex, and ancestry-matched clozapine-tolerant control was recruited  
94 from the same site. Genetic ancestry of controls and cases was further assessed by principal component  
95 analysis using the smartpca module of the EIGENSOFT package (Supplementary Figure S1).  
96 Individuals were eligible for inclusion into the study if they: (a) were aged 18 to 65 years, (b) had a  
97 clinical diagnosis of a schizophrenia-spectrum disorder, (c) had accessible medical records, and (d) had  
98 a history of clozapine therapy. For cases, an additional inclusion criterion was a history of myocarditis  
99 evidenced by abnormal troponin I/T levels ( $>2$  times upper limit of normal) or signs of left ventricular  
100 dysfunction with or without elevated C-reactive protein (CRP) levels ( $>100$ mg/L) during the first 30  
101 days of clozapine therapy, which aligns with current monitoring guidelines<sup>10</sup>. An additional inclusion  
102 criterion for clozapine-tolerant controls was evidence of clozapine therapy for a minimum of 45 days  
103 with no documented history or clinical suspicion of myocarditis. The present study was approved by  
104 the Melbourne Health Human Research Ethics Committee and University of Calgary Conjoint Health  
105 and Research Ethics Board. The study also complied with the Declaration of Helsinki and its  
106 subsequent revisions. All participants or next of kin provided written informed consent prior to  
107 participation.

108

### 109 **Whole genome sequencing**

110 Genomic DNA (750 – 1000 ng,  $A_{260}/A_{280}$  ratio  $> 1.8$ ) was isolated from whole blood samples using  
111 standard procedures and shipped to the Centre for Applied Genomics at the Hospital for Sick Children  
112 (Toronto, Canada) for library preparation and sequencing. Libraries were prepared for all samples  
113 using the Illumina TruSeq PCR-free DNA library preparation kit. Each library was subjected to 150  
114 base, paired end sequencing on three lanes of the Illumina HiSeqX.

115 The Illumina DRAGEN (version 07.021.382.3.4.9, Dynamic Read Analysis for GENomics)  
116 Bio-IT platform was used for aligning reads to the human genome (reference build hg19). After

117 alignment, sorting and duplicate marking was performed to generate aligned BAM files. Next, sample-  
118 wise gVCF files were generated using DRAGEN's haplotype calling algorithm and then these files  
119 were run through the joint calling algorithm to generate a joint VCF file. dbSNP (v150) was used to  
120 annotate variants. The jointly called VCF file was further recalibrated using the Variant Quality Score  
121 Recalibration (VQSR) algorithm to produce a high-quality variant dataset. The VQSR model was  
122 trained for single nucleotide polymorphisms (SNPs) and insertion/deletions (indels) using high quality  
123 variant data from different resources (see Supplementary Methods for more details). For SNPs, the  
124 VQSR model was trained using data from HapMap3.3, 1000 Genomes Project (Phase 1 SNP sites and  
125 Omni 2.5 M SNP array) and dbSNP (v150), with a sensitivity threshold of 99.5%. For indels, the  
126 VQSR was trained using data in the 1000 Genomes Project (gold standard sites) and Illumina Platinum  
127 sites, with a sensitivity threshold of 99.0%. Only "PASS" calls were annotated using ANNOVAR after  
128 VQSR implementation.

### 130 **Pharmacogenetic variant analysis**

131 The Stargazer v1.08<sup>11</sup> tool was used to call diplotypes, calculate activity scores, and assign genotype-  
132 inferred metabolizer phenotypes for seven genes (*CYP1A2*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*,  
133 *CYP3A5*, *CYP2E1*) relevant to clozapine metabolism<sup>12</sup>. Stargazer uses whole genome sequencing data  
134 to find phased haplotypes comprised of SNPs, small indels, and large structural variants, and then  
135 maps these haplotypes to star alleles based on translation tables maintained by the Pharmacogene  
136 Variation Consortium<sup>13</sup>. Notably, previous work has shown Stargazer recalled 100% of star alleles  
137 present in the Genetic Testing Reference Material Coordination Program's consensus genotypes<sup>11</sup>.  
138 Furthermore, a comparison of Stargazer-predicted diplotypes for the seven target pharmacogenes with  
139 the diplotypes generated by the PharmacoScan Solution (ThermoFisher, Carlsbad, USA) array in a  
140 subset of our sample (n = 22) tested on both platforms found high diplotype (94.8%) and phenotype  
141 (98.1%) concordance (Supplementary Table S1).

142 Prior to analysis, activity scores for the seven target pharmacogenes were corrected for  
143 concomitant inhibitors or inducers (Supplementary Table S2) using a method previously described<sup>14</sup>.  
144 In brief, activity scores for one or more of the seven genes were multiplied by zero if a strong inhibitor  
145 of the corresponding enzyme was present. For moderate inhibitors the activity score was multiplied by  
146 0.5 and for inducers the activity score was multiplied by 1.5. In cases where both an inhibitor and  
147 inducer of the same enzyme were present, the uncorrected activity score for the corresponding gene

148 was retained, as consensus on how to correct activity scores in this situation has not been established.  
149 The corrected activity scores were then used to assign a corrected metabolizer phenotype for each of  
150 the seven target genes. To determine differences in activity scores and metabolizer phenotypes  
151 between the cases and controls, we used the Mann-Whitney U and chi-square tests, respectively. P  
152 values were corrected for multiple testing using the Benjamini–Hochberg (B–H) step-up procedure<sup>15</sup>.

153

### 154 **Copy number variant (CNV) analysis**

155 CNVs (variants with length  $\geq 1$ kb) were called using the DRAGEN algorithm and self-normalization  
156 method. A merged CNV file was created by compiling copy number state and segmentation mean  
157 values for all samples. Segmentation mean values were log transformed to construct copy number  
158 variable regions (CNVRs). Ten regions where segmentation mean and log transformed segmentation  
159 values had conflicting CNV status were excluded (Supplementary Table S3). We used two different  
160 tools - CNVRuler<sup>16</sup> and CoNVaQ<sup>17</sup> to construct CNVRs and perform CNVR association analyses.  
161 Using CNVRuler, logistic regression was performed using the first two principal components as  
162 covariates, and a minor allele frequency threshold of greater than 0.1. CNVRuler performs multiple  
163 testing using the B–H step-up procedure<sup>15</sup>. CNVR association analysis was also performed using  
164 CoNVaQ with permutation (n=4000) testing by swapping phenotype labels. CNVR encompassing  
165 genes were retrieved from the UCSC Genome Browser.

166

### 167 **Gene burden analysis and functional annotation of candidate genes**

168 We performed burden analysis using rare and deleterious variants, defined based on the minor allele  
169 frequency from global databases and combined annotation dependent depletion (CADD) annotation  
170 score using ANNOVAR software<sup>18</sup>. A variant was considered rare if the maximum allele frequency  
171 was less than 0.05 in the 1000 Genomes Project<sup>19</sup>, ESP6500<sup>20</sup>, ExAC<sup>21</sup>, CG46<sup>22</sup> or gnomAD  
172 databases<sup>23</sup>. A deleterious variant was defined by a CADD Phred score greater than 20. We used  
173 CADD scores to include functionally important variants from protein-coding and non-coding regions.  
174 Variants that met both criteria were extracted using BCFtools (v1.9)<sup>24</sup> and burden analysis was  
175 performed on these selected variants. Two different methods were used - Combined and  
176 Multivariate Collapsing (CMC) and the sequencing kernel association test (SKAT-O), both  
177 implemented in RVTESTS<sup>25</sup>. Permutation analysis (10,000 iterations) was carried out in R by  
178 permuting phenotype labels (case and control) of observations (i.e. total allele counts) to derive

179 empirical p-values. We also used TRAPD (Testing Rare vAriants using Public Data)<sup>26</sup> software to  
180 compute the frequency of rare and deleterious alleles of prioritised genes using genome data from  
181 15,708 individuals in the gnomAD v2 database (build hg19)<sup>23</sup> and then compared these frequencies  
182 with those observed in our cases and controls using a two proportion Z-test. P-values were corrected  
183 using the Bonferroni method.

184 To evaluate the tolerance of genes for functional genetic variation, Residual Variation  
185 Intolerance Score (RVIS) scores were fetched using the web-based tool Gene Intolerance<sup>27</sup>. RVIS  
186 scores are used to rank genes in terms of their functional tolerance<sup>28</sup>. Genes with negative scores  
187 represent less tolerance to functional variations. Gene enrichment analysis was performed using  
188 ToppFun<sup>29</sup>. Finally, we assessed the mRNA expression of candidate genes in left ventricular-septal  
189 tissue from 21 healthy adult donors using published RNASeq data<sup>30</sup>. Raw RNASeq data was retrieved  
190 from the Zenodo Data repository<sup>31</sup>. Raw read counts were normalized with DESeq2, using the median  
191 of ratios method<sup>32</sup>.

## 193 RESULTS

194  
195 A summary of the study design and key results are shown in Figure 1.

### 197 Study participants and sequencing

198 Participant characteristics and relevant clinical information for the two patient groups are shown in  
199 Table 1. Cases and controls did not differ significantly in their demographic or clinical characteristics,  
200 with the exception of titration slope. On average, cases received a faster titration than controls ( $p =$   
201  $0.006$ ). However, *post hoc* comparisons of case and control titration slopes with those recommended in  
202 titration protocols used in Canada<sup>33</sup>, Australia<sup>34</sup>, and United Kingdom<sup>35</sup> showed that neither case nor  
203 control slopes differed from these national recommendations (Figure 2). In addition, *post hoc*  
204 examination showed no difference in titration slopes by ancestry (Supplementary Figure S2).

205 One of the control samples failed initial quality assessment post-sequencing, resulting in 25  
206 cases and 24 controls available for further analysis with an average depth of reads and average read  
207 mapping across the samples of 33.65x and 92.9%, respectively (Supplementary Figure S3). Ti/Tv  
208 (transition/transversion) ratio for both raw (2.00) and filtered (2.02) calls were in the acceptable whole  
209 genome range. On average, there were 3,833,700 SNPs, 953,459 indels and 619 CNVs per individual

210 (Supplementary Table S4). The final jointly called dataset contained 16,014,407 variants (both SNPs  
211 and indels).

212

213

### 214 **Pharmacogenetic variants**

215 Observed diplotype frequencies for the seven targeted pharmacogenes, along with the corresponding  
216 uncorrected activity scores and genotype-inferred metabolizer phenotypes among cases and controls,  
217 are shown in Supplementary Table S5. No significant differences in uncorrected or corrected activity  
218 scores were found between cases and controls (Supplementary Table S6) nor were differences detected  
219 in the distribution of genotype-inferred metabolizer phenotypes (Supplementary Table S7).  
220 Furthermore, titration slopes were not correlated with activity scores for the seven targeted  
221 pharmacogenes in cases or controls (Supplementary Table S8). Importantly, *post hoc* gene burden  
222 analysis did not reveal any differences in enrichment of rare and deleterious variants in the seven genes  
223 (*CYP1A2*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, *CYP3A5*, *CYP2E1*) known to be involved in  
224 clozapine metabolism.

225

### 226 **Copy number variants**

227 We identified 30,340 CNVs across the 49 samples. Comparing the average number of CNVs between  
228 cases and controls revealed no statistically significant differences in losses (cases: mean = 498,  $\pm$   
229 standard deviation = 28; controls: 503  $\pm$  19,  $p = 0.42$ ) or gains (cases: 102  $\pm$  10 vs. controls: 117  $\pm$  8,  $p$   
230 = 0.26) (Supplementary Table S9). We also did not detect significant differences with respect to the  
231 average size of the losses (cases: 10.54  $\pm$  61.6 kb vs. controls: 10.57  $\pm$  59.5 kb,  $p = 0.97$ ) or gains  
232 (cases: mean 141.37 kb, sd = 721.9 kb; controls: mean = 134.01 kb, sd = 533.5 kb,  $p = 0.66$ )  
233 (Supplementary Table S9). The CNVRuler constructed 3,178 copy number variable regions (CNVRs)  
234 among which five were found to be significant ( $p < 0.01$ , Supplementary Table S10), while the  
235 CoNVaQ tool resulted in ten significant CNVRs ( $p < 0.01$ , Supplementary Table S11). However, no  
236 CNVR survived permutation analysis or correction for multiple testing.

237

### 238 **Rare and deleterious single nucleotide variants**

239 A total of 22,703 rare and deleterious variants from 8,442 genes (out of 22,480 genic regions scanned)  
240 met criteria for gene burden analysis. On average, cases and controls had 739  $\pm$  51, and 735  $\pm$  52 rare

241 and deleterious variants per sample, respectively ( $p = 0.779$ ). We identified 114 and 44 genes ( $p <$   
242  $0.05$ ) using CMC and SKAT-O, respectively and selected 42 common genes for prioritization analysis  
243 (Supplementary Figure S4). Out of 42 genes, 15 genes had a rare and deleterious variant in our cohort  
244 (Table 2). All genes survive permutation analysis with *MLLT6* as the top candidate ( $p = 4.2 \times 10^{-3}$ ) but  
245 no gene survived multiple testing correction. However, compared to the frequencies of rare and  
246 deleterious variants in the gnomAD cohort, cases had significantly greater frequencies for seven  
247 (*MLLT6*, *CADPS*, *TACC2*, *L3MBTL4*, *NOL8*, *FAM172A*, *AFDN*) of the 15 candidate genes, after  
248 Bonferroni correction (Table 2). Frequencies in controls did not differ from those observed in the  
249 gnomAD cohort for any of the 15 genes.

250 Gene intolerance analysis showed *CADPS*, *AFDN*, and *MLLT6* were the most intolerant to rare  
251 and deleterious variation, respectively ranking amongst the 2<sup>nd</sup>, 5<sup>th</sup>, and 6<sup>th</sup> percentile of the most  
252 intolerant genes. Gene enrichment analysis of the 15 nominally significant genes revealed no  
253 significant functional category or pathway. Likewise, gene enrichment analysis including of the 15  
254 nominally significant genes and the four candidate genes from our previous GWAS (i.e. *SIX3*, *EFHC1*,  
255 *ADAM7*, *GNAI5*) revealed no significant functional category or pathway. However, all 15 genes, with  
256 the exception of *GPR179* and *SLC6A11*, were expressed in left ventricular septal tissue from healthy  
257 human adult donors with *TACC2* and *MLLT6* having the highest expression (Figure 3). Among the  
258 GWAS candidates, only *SIX3*, *EFHC1*, and *GNAI5* were expressed in left ventricular septal tissue  
259 (Supplementary Figure S5).

260

## 261 **DISCUSSION**

262 Using a whole-genome sequencing approach, our study sought to identify associations between three  
263 different and unexplored classes of DNA variation (pharmacogenetic, CNVs, rare variants) and  
264 clozapine-induced myocarditis<sup>8</sup>. Although we did not identify a single robust genetic marker for  
265 differentiating cases from controls, we identified 15 novel genes nominally associated with this  
266 important and sometimes fatal ADR - which continues to limit the use of clozapine, a highly effective  
267 medication for schizophrenia. Among the novel candidate genes identified, 13 were found to be  
268 expressed in human myocardium, specifically the ventricular septum, providing biological plausibility  
269 of a link between these genes and clozapine-induced myocarditis. Although the mechanisms by which  
270 these genes may increase risk for clozapine-induced myocarditis remain unclear and require further  
271 study, including replication in a separate patient cohort, the genes identified now represent novel

272 targets to inform future functional and clinical studies. Such studies may subsequently improve our  
273 understanding of the role played by genetic risk of this serious ADR

274 Our findings add to the list of candidate genes identified in our previous GWAS (*GNAI5*,  
275 *ADAM7*, *SIX3*, *EFHC1*) and HLA (*HLA-C\*07:01*) studies of clozapine-induced myocarditis<sup>8</sup> and  
276 suggests that the genetic architecture of this serious ADR is complex and likely polygenic. In fact,  
277 each of the candidate genes identified to date have low sensitivity and positive predictive value,  
278 limiting their clinical utility. This notion aligns with genomic studies of clozapine-induced  
279 agranulocytosis, another rare clozapine ADR, that has been associated with variants from several genes  
280 with modest sensitivity (range: 11% - 36%)<sup>36</sup>. This current state of the evidence coupled with low  
281 incidence of clozapine-induced myocarditis highlights the need for international collaboration to  
282 facilitate the rapid ascertainment of cases that will ultimately provide the statistical power required to  
283 unravel its complex genetic etiology. Recognizing this need, we formed the PROCLAIM Consortium  
284 in 2017 that currently comprises 15 recruitment sites across four countries (Canada, USA, Australia,  
285 Turkey)<sup>9</sup> and are actively looking for new sites to join the consortium.

286 Our gene enrichment analysis was inconclusive and *post hoc* exploration of the 15 genes using  
287 the published literature showed limited to no associations with phenotypes related to myocarditis,  
288 inflammation, or adverse drug reactions with few exceptions (see Supplementary Discussion).  
289 However, this is not unsurprising given the lack of previously published genetic studies examining this  
290 ADR. One notable exception was neuropeptide Y (NPY), which has a well-established role in  
291 modulating cardiac contractibility, protein degradation, and proliferation of cardiomyocytes and has  
292 been implicated in the pathogenesis of various cardiovascular morbidities (e.g., cardiac hypertrophy,  
293 arrhythmias, heart failure)<sup>37</sup>. Furthermore, clinical associations between *NPY* genetic variation and  
294 antipsychotic-induced weight gain have been reported<sup>38</sup>. Weight gain and related increases in body  
295 mass index (BMI) related to antipsychotic treatment could be clinically relevant given the number of  
296 antipsychotic trials patients undergo prior to commencing clozapine and previous work showing high  
297 BMI increases clozapine-induced myocarditis fatal outcomes<sup>6, 39</sup>. Notably, BMI did not statistically  
298 differ between our cases and controls, although there was a trend ( $p = 0.107$ , Cohen's  $d = 0.42$ ) toward  
299 higher BMIs in cases. Given the paucity of genetic markers for clozapine-induced myocarditis, future  
300 investigation of rare genetic variation in *NPY* and the other 14 putative candidate genes should be  
301 considered.

302 Our findings should be interpreted in the context of notable caveats. Since clozapine-induced  
303 myocarditis is a relatively rare event - even despite our international consortium - our patient sample  
304 size is small and had suboptimal statistical power to detect small-moderate effect sizes. This was  
305 partially mitigated by our demographically-matched case-control design and the deployment of an  
306 analysis plan appropriate for small sample sizes. Nevertheless, our results are preliminary and will  
307 require independent replication before firm conclusions can be drawn. In addition, functional  
308 characterization of the identified genes was limited to existing gene expression data derived from  
309 healthy adult human left ventricular septal tissue due to the limited biological material and clinical data  
310 available from our study participants. Notably, the PROCLAIM Consortium is currently developing a  
311 novel *in vitro* model using patient-derived induced pluripotent stem cells differentiated into beating  
312 cardiomyocytes to understand how clozapine induces myocardial inflammation and to investigate the  
313 impact common and rare genetic variants and/or epigenetic modifications have on cardiomyocyte  
314 structure and function when exposed to clozapine.

315 In summary, our whole-genome sequencing analysis identified 15 genes nominally associated  
316 with clozapine-induced myocarditis that provide novel candidates for future studies. With the possible  
317 exception of *NPY*, the mechanism by which these genes increase risk for myocarditis is unclear.  
318 Although these candidate genes will require independent validation and their roles in the development  
319 of clozapine-induced myocarditis remain to be tested, they provide potential insights into the genomic  
320 underpinnings and mechanisms by which clozapine induces myocardial inflammation and damage,  
321 that may help improve clinical management in the future.

325 **Acknowledgements**

326 The work was supported in part by the University of Calgary Cumming School of Medicine, Alberta  
327 Children's Hospital Research Institute, and University of Melbourne Establishment Grant. P.L is  
328 supported by a National Heart Foundation Future Leader Fellowship (102604).

329

330 **Competing Interests**

331 CAB is the founder and a shareholder of Sequence2Script Inc and a member of the Clinical  
332 Pharmacogenetics Implementation Consortium (CPIC) and Pharmacogene Variation Consortium  
333 (PharmVar). The remaining authors declare that the research was conducted in the absence of any  
334 commercial or financial relationships that could be construed as a potential conflict of interest.

335

336

337 **Author Contributions**

338 CB and CP conceptualized the study. AN and CB conducted the analyses and wrote the first draft of  
339 the manuscript. PL, KR, JM, MJ, NT, RS, DC, RS, SG, and CP contributed to subsequent drafts of the  
340 manuscript. All authors approved the final manuscript.

## References

1. McGorry P, Killackey, E., Lambert, T., Lambert, M., Jackson, H., Codyre, D., James, N., Pantelis, C., Pirkis, J., Jones, P., Durie, M. A., McGrath, J. J., McGlashan, T., Malla, A., Farhall, J., Herman, H. and Hocking, B. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the treatment of schizophrenia and related disorders. *The Australian and New Zealand journal of psychiatry* 2005; **39**(1-2): 1-30.
2. NICE (2014). Psychosis and schizophrenia in adults: treatment and management. .
3. Siskind D, McCartney L, Goldschlager R, Kisely S. Clozapine v. first- and second-generation antipsychotics in treatment-refractory schizophrenia: systematic review and meta-analysis. *Br J Psychiatry* 2016; **209**(5): 385-392.
4. Kilian JG, Kerr K, Lawrence C, Celermajer DS. Myocarditis and cardiomyopathy associated with clozapine. *Lancet* 1999; **354**(9193): 1841-1845.
5. Ronaldson KJ, Fitzgerald PB, McNeil JJ. Clozapine-induced myocarditis, a widely overlooked adverse reaction. *Acta Psychiatrica Scandinavica* 2015; **132**(4): 231-240.
6. Ronaldson KJ, Fitzgerald PB, Taylor AJ, Topliss DJ, Wolfe R, McNeil JJ. Rapid clozapine dose titration and concomitant sodium valproate increase the risk of myocarditis with clozapine: A case-control study. *Schizophrenia research* 2012; **141**(2-3): 173-178.
7. Youssef DL, Narayanan P, Gill N. Incidence and risk factors for clozapine-induced myocarditis and cardiomyopathy at a regional mental health service in Australia. *Australas Psychiatry* 2016; **24**(2): 176-180.
8. Lacaze P, Ronaldson KJ, Zhang EJ, Alfirevic A, Shah H, Newman L, *et al*. Genetic associations with clozapine-induced myocarditis in patients with schizophrenia. *Transl Psychiatry* 2020; **10**(1): 37.
9. Bousman C, Greenway S, Tarailo-Graovac M, Long Q, Sellmer R, Crockford D, *et al*. The Pharmacogenomics of Clozapine-Induced myocarditis (PROCLAIM) Consortium. . *Neuropsychopharmacology* 2018; S97–S98.
10. Ronaldson KJ, Fitzgerald PB, Taylor AJ, Topliss DJ, McNeil JJ. A new monitoring protocol for clozapine-induced myocarditis based on an analysis of 75 cases and 94 controls. *Aust Nz J Psychiat* 2011; **45**(6): 458-465.
11. Lee SB, Wheeler MM, Thummel KE, Nickerson DA. Calling Star Alleles With Stargazer in 28 Pharmacogenes With Whole Genome Sequences. *Clin Pharmacol Ther* 2019; **106**(6): 1328-1337.
12. Thorn CF, Müller DJ, Altman RB, Klein TE. PharmGKB summary: clozapine pathway, pharmacokinetics. *Pharmacogenet Genomics* 2018; **28**(9): 214-222.
13. Gaedigk A, Ingelman-Sundberg M, Miller NA, Leeder JS, Whirl-Carrillo M, Klein TE, *et al*. The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clin Pharmacol Ther* 2018; **103**(3): 399-401.

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

14. Lesche D, Mostafa S, Everall I, Pantelis C, Bousman CA. Impact of CYP1A2, CYP2C19, and CYP2D6 genotype- and phenoconversion-predicted enzyme activity on clozapine exposure and symptom severity. *Pharmacogenomics J* 2020; **20**(2): 192-201.
15. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Society Series B (Methodol)* 1995; **57**: 289-300.
16. Kim JH, Hu HJ, Yim SH, Bae JS, Kim SY, Chung YJ. CNVRuler: a copy number variation-based case-control association analysis tool. *Bioinformatics* 2012; **28**(13): 1790-1792.
17. Larsen SJ, do Canto LM, Rogatto SR, Baumbach J. CoNVaQ: a web tool for copy number variation-based association studies. *BMC Genomics* 2018; **19**(1): 369.
18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**(16): e164.
19. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, *et al.* A global reference for human genetic variation. *Nature* 2015; **526**(7571): 68-74.
20. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 2013; **493**(7431): 216-220.
21. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, *et al.* The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res* 2017; **45**(D1): D840-D845.
22. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010; **327**(5961): 78-81.
23. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; **581**(7809): 434-443.
24. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011; **27**(21): 2987-2993.
25. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 2016; **32**(9): 1423-1426.
26. Guo MH, Plummer L, Chan YM, Hirschhorn JN, Lippincott MF. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. *Am J Hum Genet* 2018; **103**(4): 522-534.
27. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* 2013; **9**(8): e1003709.

- 436 28. Petrovski S, Gussow AB, Wang Q, Halvorsen M, Han Y, Weir WH, *et al*. The Intolerance of Regulatory  
437 Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. *PLoS Genet* 2015; **11**(9): e1005492.  
438
- 439 29. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and  
440 candidate gene prioritization. *Nucleic Acids Res* 2009; **37**(Web Server issue): W305-311.  
441
- 442 30. Hahn V, Knutsdottir H, Luo X, Bedi K, Margulies K, Haldar S, ., *et al*. Myocardial gene expression  
443 signatures in human heart failure with preserved ejection fraction. *Zenodo Data Repository* 2020.  
444
- 445 31. Hahn V, Knutsdottir H, Luo X, Bedi K, Margulies K, Haldar S, ., *et al* (2020). Myocardial gene expression  
446 signatures in human heart failure with preserved ejection fraction: Zenodo Data Repository.  
447
- 448 32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data  
449 with DESeq2. *Genome Biol* 2014; **15**(12): 550.  
450
- 451 33. Health Canada (2020). Clozapine Product Monograph. Health Canada: Ottawa.  
452
- 453 34. ACSQHC (2012). National Adult Clozapine Titration Chart User Guide. Australian Commission on Safety  
454 and Quality in Health Care Sydney.  
455
- 456 35. NHS (2019). CLOZAPINE TREATMENT GUIDELINES. National Health Service Foundation Trust: United  
457 Kingdom.  
458
- 459 36. Legge SE, Walters JT. Genetics of clozapine-associated neutropenia: recent advances, challenges and  
460 future perspective. *Pharmacogenomics* 2019; **20**(4): 279-290.  
461
- 462 37. Tan CMJ, Green P, Tapoulal N, Lewandowski AJ, Leeson P, Herring N. The Role of Neuropeptide Y in  
463 Cardiovascular Health and Disease. *Front Physiol* 2018; **9**: 1281.  
464
- 465 38. Shams TA, Müller DJ. Antipsychotic induced weight gain: genetics, epigenetics, and biomarkers  
466 reviewed. *Curr Psychiatry Rep* 2014; **16**(10): 473.  
467
- 468 39. Ronaldson KJ, Fitzgerald PB, Taylor AJ, Topliss DJ, McNeil JJ. Clinical course and analysis of ten fatal  
469 cases of clozapine-induced myocarditis and comparison with 66 surviving cases. *Schizophrenia*  
470 *research* 2011; **128**(1-3): 161-165.  
471  
472  
473

474 **Figure Legends**

475  
476 **Figure 1.** PROCLAIM study design, analysis and key results. Age, sex and ethnicity matched  
477 treatment resistant schizophrenia patients who did (cases) or did not (controls) develop myocarditis  
478 following clozapine exposure were recruited. Using whole genome sequencing data, candidate  
479 pharmacogenetic variants, copy number variable regions and rare variants were examined for their  
480 association with clozapine-induced myocarditis.

481  
482 **Figure 2.** Distribution of titration slopes by history of clozapine-induced myocarditis. Cases differed  
483 significantly from controls ( $p = 0.006$ ). Dashed lines represent titration slopes recommended by  
484 clozapine dosing protocols in Canada, United Kingdom, and Australia. Case titration slopes did not  
485 differ from slopes recommended by protocols used in Canada ( $p = 0.544$ , United Kingdom ( $p = 0.795$ ),  
486 or Australia ( $p = 0.259$ ). Likewise, control titration slopes did not differ from slopes recommended by  
487 protocols used in Canada ( $p = 0.187$ , United Kingdom ( $p = 0.248$ ), or Australia ( $p = 0.869$ ).

488  
489 **Figure 3.** Normalised RNASeq read counts of the 15 nominally associated loci identified in gene  
490 burden analysis in left ventricular septal tissue of healthy adult donors ( $n=21$ ). *MLLT6*: PHD Finger  
491 Containing, *CADPS*: Calcium Dependent Secretion Activator, *TACC2*: Transforming Acidic Coiled-  
492 Coil Containing Protein 2, *L3MBTL4*: L3MBTL Histone Methyl-Lysine Binding Protein 4, *NPY*:  
493 Neuropeptide Y, *SLC25A21*: Solute Carrier Family 25 Member 21, *PARVB*: Parvin Beta, *GPR179*: G  
494 Protein-Coupled Receptor 179, *ACAD9*: Acyl-CoA Dehydrogenase Family Member 9, *NOL8*:  
495 Nucleolar Protein 8, *NADK2*: NAD Kinase 2, Mitochondrial, *FAM172A*: Family With Sequence  
496 Similarity 172 Member A, *AFDN*: Afadin, Adherens Junction Formation Factor, *SLC6A11*: Solute  
497 Carrier Family 6 Member 11, *PXDN*: Peroxidasin

**Table 1.** Demographic and clinical characteristics of participants with (N = 25) and without (N = 25) a history of clozapine-induced myocarditis

Characteristic	Clozapine-induced myocarditis	Clozapine-tolerant controls	P
Age, mean (sd) years	36 (13)	35 (11)	0.650
Sex, n (%) males	18 (72%)	19 (76%)	0.747
Self-Report Ancestry <sup>a</sup> , n (%)			0.695
European	21 (84%)	20 (80%)	
East Asian	2 (8%)	3 (12%)	
South Asian	0 (0%)	1 (4%)	
Near Eastern	1 (4%)	1 (4%)	
Oceanian	1 (4%)	0 (0%)	
Body mass index, mean (sd)	31 (10)	27 (9)	0.107
Highest dose received in first 28 days, mean (sd)	274 (84)	310 (150)	0.366
Titration slope <sup>b</sup> , mean (sd)	20 (8)	14 (6)	<b>0.006</b>
Days from clozapine initiation to myocarditis, mean (sd)	15 (5)	–	–
Valproic acid use, n (%)	6 (25%)	8 (32%)	0.588
Concomitantly used inhibitors, n (%)			
CYP1A2 inhibitor <sup>c</sup>	0 (0%)	0 (0%)	1.000
CYP2C9 inhibitor <sup>c</sup>	0 (0%)	0 (0%)	1.000
CYP2C19 inhibitor <sup>c</sup>	2 (8%)	2 (8%)	1.000
CYP2D6 inhibitor <sup>c</sup>	5 (20%)	6 (24%)	0.733
CYP2E1 inhibitor <sup>c</sup>	0 (0%)	0 (0%)	1.000
CYP3A4/5 inhibitor <sup>c</sup>	0 (0%)	0 (0%)	1.000
Concomitantly used inducers, n (%)			
Tobacco (smoked)	6 (26%)	10 (40%)	0.307
Carbamazepine	0 (0%)	1 (4%)	0.312

<sup>a</sup> Ancestry groupings are based on the biogeographic grouping system described by Huddart et al., Clin Pharmacol Ther. 2019, 105(5): 1256-1262.

<sup>b</sup> Slope = (highest dose in first 28 days - starting dose) / (days of titration - 1). A higher slope translates to a faster titration.

<sup>c</sup> Medications taken by participants that are known to be moderate or strong inhibitors of CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP2E1 or CYP3A4/5 according to the Flockhart Drug Interactions Table (<https://drug-interactions.medicines.uq.edu.au/Main-Table.aspx>)

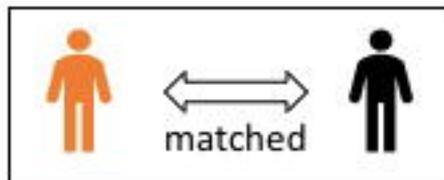
**Table 2.** Top hits from gene burden analysis

Genes	Frequency Controls (n=24)	Frequency Cases (n=25)	Frequency gnomAD (n=15,708)	Case vs Control			Two Proportion Z-test		RVIS (Percentile) <sup>1</sup>
				CMC P-value	SKAT-O P-value	Empirical P-value	Case vs gnomAD P-value	Control vs gnomAD P-value	
MLLT6	0.000	0.140	0.007	0.0051	0.0324	0.0042	3.9x10 <sup>-15</sup> *	0.6818	-1.51 (6.0%)
CADPS	0.000	0.120	0.025	0.0104	0.0207	0.0101	0.0024*	0.4354	-2.30 (2.2%)
TACC2	0.021	0.160	0.024	0.0119	0.0500	0.0103	9.9x10 <sup>-6</sup> *	0.9203	1.92 (96.5%)
L3MBTL4	0.000	0.100	0.007	0.0208	0.0354	0.0198	3.5x10 <sup>-8</sup> *	0.6818	0.38 (66.9%)
NPY	0.000	0.100	0.032	0.0208	0.0208	0.0208	0.0539	0.3735	-0.08 (45.5%)
SLC25A21	0.021	0.140	0.037	0.0240	0.0344	0.0209	0.0065	0.6745	0.51 (72.2%)
PARVB	0.000	0.100	0.025	0.0208	0.0445	0.0211	0.0166	0.4354	0.29 (62.8%)
GPR179	0.000	0.100	0.039	0.0208	0.0399	0.0218	0.1158	0.3222	2.80 (98.9%)
ACAD9	0.000	0.100	0.028	0.0208	0.0491	0.0218	0.0295	0.4056	-0.69 (21.3%)
NOL8	0.000	0.100	0.049	0.0208	0.0491	0.0221	0.2382	0.2670	1.92 (96.5%)
NADK2	0.000	0.100	0.015	0.0208	0.0491	0.0222	0.0005*	0.5485	0.30 (63.4%)
FAM172A	0.021	0.140	0.016	0.0240	0.0457	0.0225	9.2x10 <sup>-7</sup> *	0.8145	-0.38 (32.2%)
AFDN	0.000	0.100	0.017	0.0208	0.0208	0.0227	0.0013*	0.5222	-1.65 (5.0%)
SLC6A11	0.000	0.120	0.029	0.0208	0.0292	0.0239	0.0068	0.3953	-0.71 (20.7%)
PXDN	0.000	0.100	0.009	0.0208	0.0399	0.0253	0.0348	0.6384	-1.08 (11.4%)

507 Comparing frequency of qualifying rare and deleterious variants per gene between cases and controls in the PROCLAIM cohort.  
 508 Frequencies of the qualifying rare variants in cases and controls were also compared with public controls (gnomAD v2 public  
 509 database).  
 510 \*significant after Bonferroni correction (0.05/15 tests = 0.0033). 1 = RVIS Score, a gene-based gene intolerance score based on ExAC  
 511 v2 release. Genes with negative scores are more “intolerant” than those with positive scores. *MLLT6*: PHD Finger Containing,  
 512 *CADPS*: Calcium Dependent Secretion Activator, *TACC2*: Transforming Acidic Coiled-Coil Containing Protein 2, *L3MBTL4*:  
 513 L3MBTL Histone Methyl-Lysine Binding Protein 4, *NPY*: Neuropeptide Y, *SLC25A21*: Solute Carrier Family 25 Member 21,  
 514 *PARVB*: Parvin Beta, *GPR179*: G Protein-Coupled Receptor 179, *ACAD9*: Acyl-CoA Dehydrogenase Family Member 9, *NOL8*:  
 515 Nucleolar Protein 8, *NADK2*: NAD Kinase 2, Mitochondrial, *FAM172A*: Family With Sequence Similarity 172 Member A, *AFDN*:  
 516 Afadin, Adherens Junction Formation Factor, *SLC6A11*: Solute Carrier Family 6 Member 11, *PXDN*: Peroxidas

517  
518

Clozapine-Induced  
Myocarditis  
(n = 25)



Clozapine-Tolerant  
Controls  
(n = 25)



Whole genome Sequencing  
(Illumina HiSeq, 100 bp paired, Average depth 30x)

Pharmacogenetic Variants

Metabolizer  
phenotypes

No significant results  
*Tables S5 and S6*

Copy Number Variations

Association  
analysis

No significant results  
*Tables S10 and S11*

Rare Variants

Gene Burden  
Analysis

15 nominally significant  
*Table 2*

Titration Slope

