The Healthy Nevada Project: rapid recruitment for population health study

Joseph J Grzymski[*1,2] Max J Coppes [2,3,4] Jim Metcalf [1,2] Christos Galanopoulos [2,3] Chris Rowan [2,3] Michele Henderson [2,3] Robert Read[1] Harry Reed [1] Bruce Lipp [1] Dave Miceli [1], Susan Rybarski [1], Anthony Slonim[3*].

1. Desert Research Institute
    Applied Innovation Center
    2215 Raggio Parkway
    Reno, NV 89512
2. Renown Institute of Health Innovation
    2215 Raggio Parkway
    Reno, NV 89512
3. Renown Health
    50. W Liberty St
    Reno, NV   89501
4. University of Nevada Reno School of Medicine
    1664 N Virginia Street
    Reno, NV, 89557

*Corresponding authors: joeg@dri.edu, aslonim@renown.org

28 **Abstract**

29 Background

30   Nevada ranks in the bottom half of overall health rankings in the United States. The majority of

31 residents of Northern Nevada live in Washoe County, which is confounded with high age-adjusted death

32 rates for heart disease, cancer and chronic lower respiratory disease.

33 Methods

34   Saliva as a source of DNA was collected from adults in Northern Nevada as the first phase of a

35 much larger (100,000 participants) effort to contribute to comprehensive population health studies in

36 Nevada. The personal genetics company 23andMe was used to genotype the first 10,250 participants

37 and deliver their custom ancestry, traits, wellness, and carrier status reports.

38 Results

39   The study was announced by Governor Brian Sandoval on September 15, 2016 and within two

40 days the registration of 9,700 volunteers for an appointment was complete. Processing of 9,344

41 participants was achieved in 3 months, with a no-show rate of just over 11%. The participant population

42 was skewed to female and was less racially diverse than the population.

43 Conclusion

44   DNA genotyping was administered free-of-charge and the patient population was representative

45 of the socio-economic diversity in northern Nevada – indicating that free genetic testing is of interest to

46 a broad swath of the population and a powerful motivator for comprehensive population health study

47 research.

48
49
50 **Introduction**

51 Health is the result of the complex and not fully elucidated interaction between genetic,

52 sociodemographic and environmental factors often influenced by the availability or quality of

53 health care [1]. Optimally, these factors need to be integrated in order to develop and apply

54 interventions to improve the health of a population.

55

56 Depending on the data source used, Nevada ranks in the bottom half of overall health rankings

57 in the United States. The 2016 annual report of America's Health Ranking for example, ranks

58 Nevada #35 [2]. At the same time, healthcare spending per capita in Nevada is $5,735, 16% below

59 the national average, ranking the state #45 [3]. The majority of residents of Northern Nevada

60   live in Washoe County, which is confounded with high age-adjusted death rates for heart disease,

61   cancer and chronic lower respiratory disease. These death rates exceed those in Nevada and the

62   United States more generally (cumulative deaths in the 3 categories= 501.5 in Washoe County

63   vs. 416.5 in Nevada vs. 385 in the United States per 100,000) [4].

64

65   The Renown Institute for Health Innovation (IHI), a newly established public-private partnership

66   between Renown Health (RH) and the Desert Research Institute's (DRI) Applied Innovation

67   Center (AIC), aims to better understand those factors that contribute to poorer health outcomes

68   in Nevada. This will be achieved by identifying and prioritizing investigations that will lead to

69   insights allowing Nevadans to develop strategies that will improve overall health. The Healthy

70   Nevada Project (HNP) was aimed at two goals. First, to start informing individual Nevadans of

71   their genetic makeup such that they might make better informed decisions to modify their

72   behaviors and improve their health. Since genetic testing as an educational tool is likely biased

73   and restricted to those who can afford the tests (in general selecting for a more affluent

74   population), ours were provided at no cost to participants. Second, to collect comprehensive

75   genetic data on an extensive number of Nevadans that could be used to cross-reference with

76   health, healthcare, social and environmental data to identify specific associations in our local

77   communities and assist in healthcare resource planning. This will be achieved using our expertise

78   in geomapping and understanding the effects of natural and human-induced environmental

79   changes, and the unique computational and storage demands for data hosting, exploration, and

80   visualization. The latter will be used to analyze aggregated electronic health record (EHR) data

81   provided by RH.

82

83   **Material and Methods**

84   Patient Population

85   According to the United States Census data for 2016, the Washoe County catchment area, at the

86   center of this phase of the study, includes 6,302 square miles and is home to more than 453,000

87   people (population per square mile 71.9). Of these 22.2% are under the age of 18 years and were

88   excluded from this, first, phase of the study. The study was open to all other members of the

89   community, including, the approximately 13.8% of the population that lives below the Federal

90   Poverty Level.

91

92   <u>Study Appointment Scheduling</u>

93   23andMe has in place a well-established registration process where individuals can send their

94   saliva from the privacy of their home. However, for the purpose of this study, which included an

95   additional consent to merge de-identified genotype data with de-identified EHR data, an

96   additional registration was required. It was also important to educate each participant

97   thoroughly on the study, its implications, and potential outcomes for Nevada. For these reasons,

98   it was decided to provide participants the opportunity to meet research staff face to face at a

99   physical location for a study participation experience that was well choreographed. This process

100   included a study "meet up" location, a chaperone to the study location, and a video tour of the

101   process. These methods are detailed below.

102

103   Investigators utilized Appointment-Plus, a 3rd party HIPAA-compliant commercial appointment

104   provider service [5]. The software allowed study coordinators to manage subject load balance

105   and modify appointment schedules given regular no shows and cancellations.

106

107   To schedule an appointment, prospective study participants were guided to a website where they

108   entered basic participant information, including name, birth date, and email address, and

109   subsequently selected a one hour-long time slot to meet research staff. Typically, five, one-hour

110   time slots were offered during each weekday, two, one-hour time slots each Saturday, and one,

111   one-hour time slot each evening from Monday to Friday. Each time slot could include up to 30

112   participants.

113

114   After scheduling an appointment, research coordinators validated the appointment, after which

115   participants were sent an email through Appointment-Plus confirming their registration along

116   with directions to the collection site. In addition, the confirmation email contained the RH-DRI

117   Community Health Project consent document in a Portable Document Format (PDF).

118   Appointment-Plus generated a reminder for the appointment, emailed two days prior to the

119   scheduled visit.

120

121   <u>Institutional Review Board (IRB) and Consent</u>

122   All participants were provided with 3 consents. Two consents were provided by 23andMe in

123   electronic format, one to consent for genetic testing and sharing genotype data with the study,

124   the other for DNA bio-banking. The latter consent is optional and does not affect the 23andMe

125   experience or access to genotype data. The third consent form was provided by IHI and provided

126   permission for the de-identified genetic data to be linked to de-identified EHR data from the

127   same study subjects.

128   *23andMe Consent*

129   At the appointment, participants registered with 23andMe via laptop computers available at the

130   sampling site. During this online registration, participants provided 23andMe electronically with

131   the required consents, similar to any other individual who chooses to have their DNA tested by

132   23andMe. Participants approved the anonymous utilization of their SNP data and also optionally

133   approved for the saliva to be bio banked. Both 23andMe consent forms have been approved by

134   the Ethical and Review Services IRB. Registration with 23andMe is the point where the unique

135   saliva collection tube number is tied to subject demographic information and an email address

136   for 23andMe account creation and further communication.

137   *IHI Consent*

138   A copy of the IHI study consent document approved by the University of Nevada Reno IRB was

139   provided electronically to each participant when they received an electronic confirmation of their

140   registration. This gave each participant the opportunity to study the IHI study consent form and

141   prepare questions when they met with study coordinators. During the one-hour appointment,

142   special effort was made to ensure that participants understood the data sharing authorization

143   granted by signing the IHI consent, which would provide the IHI access to the 23andMe genotype

144   data in de-identified form. At their appointment, participants signed the IHI consent form,

145   providing the investigators a permanent record of consent. Each participant received a signed

146   copy of the IHI consent form.

5

147   A total of 270 participants were provided the option to participate without a formal appointment

148   – the majority of these people were guests on the day of the project launch and 23 were

149   "ambassadors" pre-identified to communicate about the study to the community. These

150   individuals received a consent form immediately prior to providing a saliva sample. However,

151   ample time was provided for them to read and consider the content of the consent form.

152   All participants signed the consent form with a Collaborative IRB Training Initiative (CITI) certified

153   study coordinator available to answer questions and as witness prior to sample collection.

154

155   <u>Acquisition collection kits</u>

156   23andMe shipped 10,000 saliva collection kits to Renown Regional Medical Center. These kits

157   conform to accepted FDA saliva collection requirements and contain a saliva sample tube

158   (Orogene DX OGD-500.001) with a unique identifier. Each participant is identified by an 11

159   character mixed-case string to uniquely track study participants. There is no associated protected

160   patient information associated with this participant identifier that is accessible to researchers.

161   Only the IHI Principal Investigator has the link between the participant's ID and the research data.

162

163   <u>Sampling Location</u>

164   The sampling was performed in the Renown Regional Medical Center's wing designated for

165   Research and Education. Each cohort would show up at a public gathering place in the hospital.

166   Research coordinators then checked participants in, arranged them in groups of 10, and escorted

167   them to the study area. The study area was divided into four separate stations: education,

168   23andMe registration & consent, saliva collection, and sample deposition followed by a final

169   paperwork and quality assurance check by a CITI trained coordinator.

170

171   A video was produced to ensure that participants understood and followed the registration and

172   saliva collection process. When cohort members arrived at the saliva collection site they first

173   watched a seven-minute video with instructions for 23andMe registration, and explanation of

174   the RH-DRI study, and saliva collection. The video ensured consistent information was provided

175   to the participants and provided a scalable method of communication to research participants.

176 Next, participants walked to the 23andMe registration station where their test kits were waiting

177 to be registered. There they opened the kits and registered the unique tube number with

178 23andMe and completed the standard 23andMe online consents. From there, participants went

179 to the saliva collection site to produce the requisite 2 ml sample. This area was partitioned for

180 privacy. After sample collection, participants took their sealed kit to the sample deposition

181 station where they dropped it off and had their paperwork confirmed by a CITI trained research

182 coordinator. Finally, participants were then escorted back to the gathering area.

183

184 English as a Second Language

185 Recognizing a large Hispanic community in Washoe County, a separate one-day saliva collection

186 event was conducted in Spanish. This required that the consent forms and signage be converted

187 to Spanish and approved by the IRB. Moreover, Spanish-speaking study coordinators were

188 available to coordinate the sample collection/registration in Spanish. A total of 37 Hispanic

189 research participants were recruited during this one-day event.

190

191 Collection in remote areas in Nevada

192 Nevada has remote areas that are sparsely populated and underserved medically. The desire to

193 include participants from these remote areas and to test our ability to collect at these sites, made

194 us set up offsite saliva collection events in Winnemucca, Yerington, and Fallon. Each event was

195 staffed with 6 study coordinators, who brought all necessary equipment (laptops, saliva

196 collection kits, signage, pens, tape and other supplies) with them in a van.

197

198 Cost to participants

199 Through a generous donation of the Renown Health Foundation, we were able to offer the DNA

200 testing for free to all participants. The Governor's Office of Economic Development through a

201 Knowledge Fund grant to the DRI AIC helped fund data scientists.

202

203 Staffing

204 Support for study execution depended on numerous staff from RH and from DRI. At RH nine FTEs

205 were utilized for research coordination; at DRI four FTEs were utilized to ensure patients could

206 concurrently register for 23andMe and the IHI study. DRI also spearheaded utilization of the

207 appointment system and creation of the two web-based applications for the registration process.

208

209 EHR data

210 Starting in June 2016, de-identified EHR data spanning more than 10 consecutive years of >1.4

211 million patients encompassing >26 million provider encounters were transferred to AIC's Health

212 Insurance Portability and Accountability Act (HIPAA) secure analysis and storage environment.

213 Since then, EHR data are updated quarterly.

214

215 Matching

216 In order to link subject EHR data to SNP data the participant 23andMe ID was mapped to a

217 participant's email address, name and date of birth and then de-identified.

218

219 Cardiac Risk Score Calculation.

220 In order to demonstrate the power of IHI's data, we tested, to demonstrate proof of principle,

221 our ability to replicate a recent study of genetic risk scoring for coronary artery disease using SNP

222 data [6]. Variant calls from the Human Omniexpress-24 array were provided by 23andMe as VCF

223 files. The data were imputed and phased using the Michigan Imputation Server and Eagle v2.3

224 [7,8]. The reference panel was HapMap2. Reference SNP cluster IDs (rsIDs) were associated with

225 genomic coordinates using SNPTracker [9]. For cardiac risk score calculation, a total of fifty SNPS

226 were examined based on reported findings [6]. Forty-eight of the 50 rsIDs were in our dataset;

227 31 were genotyped and 17 were imputed. A cardiac risk score was then assessed based on the

228 odds ratio of each SNP, as previously reported [6]. The natural log of the published odds ratio at

229 each SNP was multiplied by the number of risk alleles for each de-identified participant. A final

230 sum was then calculated across all variants. Data were then visualized using a histogram with a

231 bin width of 0.05.

232

8

233   **Results**

234   Participation and no-show rate

235   Restricting recruitment to Northern Nevada participants in this phase of our study was primarily

236   guided by geographic convenience with an estimated local population of about 450,000 people

237   in Washoe County. Although the study and announcement by the Governor was covered in local

238   news and press the appointment process was through the Renown Health website. In this early

239   stage, we wanted to demonstrate our ability to develop an infrastructure that would allow for

240   the successful collection of saliva samples from thousands of people in a short time period. Not

241   only were we able to successfully collect saliva from participants in Washoe county, but we also

242   developed an effective 'mobile sample collection' mechanism, to obtain DNA from people living

243   in communities in rural Nevada.

244

245   We were surprised at the incredible speed with which participants signed on. The study was

246   announced by Governor Brian Sandoval on September 15, 2016 and within two days the

247   registration of 9,700 volunteers for an appointment was complete. Three hundred volunteer slots

248   were reserved for special events. Moreover, ongoing interest for participation was evidenced by

249   a waiting list of more than 4,000 participants. The no-show rate was just over 11% (1,210 of

250   10,554 online appointments made) and was mitigated by pulling people from a waitlist.

251

252   Formal appointments began on September 28, 2016 and were completed on December 21, 2016.

253   Within that time frame there were 69 days of appointments, 17 days off and 9,344 appointments

254   were completed.

255

256   Planning and success

257   Considerable effort went into planning the public launch. Critical to garnering high-level support

258   in all three organizations was ensuring the Governor of Nevada was interested in participating.

259   The Governor readily agreed to be the first participant in the study on launch day. Dignitaries,

260   board members and employees of both RH and DRI were invited to attend the launch along with

261   members of the press. Additionally, 23 ambassadors were selected from the community to

9

262   highlight the study. These emissaries were carefully chosen for their visibility in our communities,

263   ranging from a Native American with a high-level position in their tribe, to a radio disc jockey who

264   talked about the study on his radio show.

265

266   Registration and email address

267   Using a personal email address was a crucial component to the study. First, 23andMe uses the

268   email address as the only mechanism to communicate results with its consumers. When

269   individuals, often a spouse, attempted to utilize a shared email address, the 23andMe

270   registration process, structured to disallow sharing email addresses, would issue a warning "This

271   email is already in use". Second, the IHI uses the email address provided to identify participants.

272   Occasionally, participants would have used an email to register with 23andMe and a different

273   one when making the appointment for saliva collection. These differences were reconciled so

274   that one email address was used in association with one individual.

275

276   Results from 23andMe

277   Within approximately eight weeks after donating a saliva sample, each subject received results

278   from 23andMe's Personal Genome Service, including more than 60 personalized genetic reports

279   that detail their carrier status, genetic mutations which can be passed onto children, for inherited

280   genetic conditions, as well as their traits and ancestry. Earlier this year, the FDA approved

281   23andMe to provide participants information on their risks to get certain conditions like

282   Parkinson and Late-onset Alzheimer disease. In doing so, 23andMe is the only company with FDA

283   approval to provide genetic health risk reports without prescription. The new information is

284   being made available to participants through 23andMe. Combining genetic research for which

285   results are not immediately available with 'immediate gratification' through the 23andMe

286   Personal Genome Service was appreciated by many participants. In fact, we suspect that many

287   participants were swayed to partake because the availability of immediate genetic results, a

288   hypothesis that we intend to test.

289

290   Epidemiology (description of basic demographic data)

10

291   *Race/ethnicity*

292   The racial and Hispanic origin of the broad study region defined as Northern Nevada is compared

293   to Washoe County, the rest of Nevada and the USA in Table 1. Northern Nevada (population

294   713,413) is comprised of 11 counties: Carson City, Churchill, Douglas, Elko, Eureka, Humboldt,

295   Lander, Lyon, Pershing, Storey, and Washoe. The majority of residents live in Washoe County

296   (population 453,000). Almost 80% of Northern Nevadans are White and 20% self identify as

297   Hispanic or Latino. Table 2 shows that the study cohort is less diverse than the reported race and

298   ethnicity distribution in Washoe County.

299

300   *Residence*

301   The residence of the study participants is reported in Supplemental Table 1. In Northern Nevada,

302   RH is the only tertiary care facility between Sacramento, California and Salt Lake City, Utah,

303   attracting patient referrals from surrounding counties. In addition, the Sierra Mountains and Lake

304   Tahoe are vacation destinations that attract a significant number of visitors. This explains the

305   discrepancies noted in Supplemental Table 1: roughly one-third of the RH patient population is

306   from outside of Northern Nevada, while over 98% of study participants are from Northern

307   Nevada. Moreover, for this initial part of the study, the majority of consents were performed at

308   the Renown Regional Medical Center in Reno, NV, self-selecting people living close to the

309   hospital. However, the three rural recruitment events in Fallon, Winnemucca, and Yerington,

310   were highly successful and demonstrated our ability to use a 'mobile sample collection' system

311   to obtain saliva for this study from people living in rural Nevada. A total of 501 participants were

312   recruited from these off-site locations (Fallon 160, Winnemucca 168, and Yerington 173). In

313   addition, 551 people from rural Northern Nevada travelled to our central collection center in

314   Reno, resulting in 1,052 participants from outside Washoe County.

315

316   *Age*

317   Age distribution for USA, Nevada, Northern Nevada, and Washoe County are provided in

318   Supplemental Table 2[10]. Of interest the lower % of people under 30 years of age in Northern

319   Nevada as a whole compared to Nevada and the United States (39% versus 41 % and 42%

320   respectively). Washoe County has 42% of people in this age group.

321   When comparing the age distribution between the Healthy Nevada Project (HNP) participants

322   and the RH EHR patient population we noted several differences (Supplemental Table 3). The

323   HNP participants age distribution is skewed to older participants because individuals 18 years

324   and younger were specifically excluded for this first phase of the study. Second, HNP participants

325   between the ages of 31 and 70 years were over-represented in this cohort, while those over the

326   age of 70 were underrepresented.

327

328   *Income/poverty level*

329   The Reno/Sparks urban area has 5 zip codes with greater socio-economic disparity than the rest

330   of Washoe County [4]. Approximately 30% of the population live in these 5 zip codes but

331   represent 42% of the hospital inpatient visits and 54% of the emergency room visits.

332   Supplemental Table 4 demonstrates that the inclusion criteria and our ability to offer the test for

333   free, allowed us to include people from various economic backgrounds. The percentage of

334   participants considered to live below the poverty level did not select only for participants who

335   could afford personal genetic testing. The participant population comprised of approximately

336   30% who live in the most impoverished Zip Codes of Washoe County indicating no cost

337   participation does provide economic diversity to the cohort.

338

339   *Gender*

340   The EHR data population was well balanced in terms of gender reflecting the gender balance in

341   the overall population. The HNP population consisted of many more females (65%) than males

342   (35%)(Supplemental Table 5).

343

344   Sample issues

345   On occasion, there were sample issues – not enough DNA from saliva or other methodological

346   issues – that caused a sample exception in the 23andMe processing pipeline. These were

347   communicated to participants and a new kit was sent to their address of choice. We encountered

348 21 sample exceptions in our study. The largest problem encountered with the delivery of results

349 to participants was forgotten passwords or other issues accessing established 23andMe

350 accounts. Participants would call the study coordinator number and be directed to 23andMe

351 customer service for privacy purposes (emails and passwords were involved) while an IHI

352 research coordinator followed up to ensure the participant was satisfied with the resolution. Five

353 people dropped out of the study because of dissatisfaction with the experience or concerns with

354 privacy.

355

356 Electronic Healthcare Records (EHR)

357 Participants consented to have their EHR history at RH, if available, merged with their genetic

358 data in order to study outcomes – this matching was done and then de-identified. Since this

359 genetic study was not restricted to just RH patients, 8180 participants were EHR matched, while

360 the remaining 1520 participants had no EHR data or were unable to be matched.

361

362 Cardiac Risk Scores

363 Cardiac risk scores were quantified from the SNP data provided by the direct to consumer genetic

364 testing company, 23andMe [6]. The mean risk score of our cohort (3.57) was consistent with

365 other studies (Figure 2). The lowest quartile was 3.208 and upper quartile was 3.918.

366

367 **Discussion**

368 The ultimate goal of the Healthy Nevada Project is to recruit 100,000 study participants,

369 representative of Nevada's population to improve the health of the population through

370 identification of susceptibility for disease, early detection, precision treatment approaches and

371 care delivery research, among many possibilities. This will allow us to link detailed health and

372 healthcare information with comprehensive environmental data and extensive genetic

373 information. Here we report on the first phase of DNA sample recruitment, the acquisition of

374 DNA samples from 9,700 individuals in Northern Nevada, and the proof of principle that our

375 approach can generate the desired data to influence health in our communities.

376

377 The enrollment was completed in two days and the consenting and saliva collection was

378 completed in 69 working days. We suggest that the remarkable success can be attributed to a

379 carefully organized marketing campaign and an extremely successful public launch. Prior to the

380 launch, with the input of a local advertising agency, the study was widely advertised ranging from

381 creating a web presence to billboards in the community. Throughout the campaign special

382 attention was given to ensuring consistent messaging. The following factors also contributed to

383 the success of the study. First, garnering sufficient organizational support to conduct the study.

384 This applies in particular to RH that had to make space and personnel available for scheduling

385 sample collections, hosting, informing, and consenting participants, and for collecting and

386 processing the samples. Second, the generous financial support of the Renown Health

387 Foundation was of critical importance since it removed financial barriers to participation. Third,

388 ensuring awareness that the study was taking place in the community and was intended to

389 address a community problem. This involved RH, the DRI, the Governor's office, and 23andMe.

390 As part of driving home the marketing message, 270 launch attendees were given 23andMe saliva

391 collection kits to begin the study after completing the consent process onsite -- these included

392 23 ambassadors who were instrumental in helping publicize and share through their networks

393 the potential benefits of participating. Finally, utilizing 23andMe's well-established processes

394 and infrastructure to facilitate sample and data collection, as well as subsequent DNA extraction

395 and processing was a key element of the study. Partnering with 23andMe, who has

396 commoditized the process, allowed for quick, efficient, and quality sample collection and

397 processing.

398

399 Privacy

400 While potential privacy concerns related to access of the combined (health, healthcare, genetic,

401 social, and environmental) data were clearly dealt with in the consent forms provided, a small

402 number of participants remained concerned with privacy issues associated with the study.

403 Specifically, some participants asked research coordinators if they were required to use their real

404 names, email addresses and birthdates. Incorrect email addresses make account creation with

405 23andMe impossible while incorrect names and/or birthdates would make matching participants

406 to EHR data impossible. We also noted that less than 10 potential participants, who indicated to

407 have intensively examined the privacy statement on the 23andMe website, opted out at the time

408 of saliva collection. Finally, we have a few participants specifically inquiring whether IHI would

409 share their DNA data with insurance companies. The IHI IRB document expressly states that only

410 DRI will have access to both de-identified DNA and EHR. Consequently, no identifiable genetic

411 data will ever be available to RH, its employees, or any other organization, including insurance

412 companies. This firewall between the hospital system and the research was considered essential

413 to the study. The only way to link the SNP data to EHR record data would be after obtaining

414 specific consent from participants for future studies that would have to be reviewed and

415 approved by an IRB.

416

417 Biases in our convenience sample

418 There are obvious biases in convenience sampling [11]. The RH population is noticeably on

419 average older; older people use hospital care more frequently than younger (explaining the

420 higher percentages in the EHR data group), while younger people are generally more mobile and

421 able to travel to the study site (explaining the higher percentages in the HNP population). When

422 expanding the IHI study to 100,000 participants special attention will need to be made to recruit

423 more aged individuals to have an accurate reflection of the population.

424

425 There are noted sex based biases in volunteer behavioral research [12]. As well, a large majority

426 of the appointments were during the day and it could be that the older population had higher

427 male:female workforce participation. Since during this first phase of the study there had been no

428 attempt to balance the male:female ratio, we have identified a need to gain that balance going

429 forward, when the study expands and will purposefully focus on a large cohort representative of

430 the general Nevada population [13,14].

431

432 Though NV is not as ethnically diverse as other states our study population was skewed toward

433 whites and under-represented the Hispanic population. Better bi-lingual advertising and direct

434 communication with the Latino community will be necessary for the future study expansion.

435

436 <u>Using consumer genetics in a population health study</u>

437 Genotype data collected from consumer platforms coupled with self-reported phenotype
438 information have great potential [15]. However, how these data are communicated and
439 implemented in a healthcare setting is still a challenge. We are in the early stages of developing
440 responsible methods to integrate these metrics into more traditional risk factors such as the
441 American Heart Association healthy lifestyle factors of no current smoking, BMI<30, at minimum
442 weekly physical activity and healthy diet [16]. Yet, it is remarkable that for minimum cost (in
443 healthcare perspectives) we can genotype at population scale and gain important insights into
444 our community genetic health risk.

445

446 <u>Lessons learned</u>

447 Part of the process was making a video available to each participant that walked him or her
448 through the process and expectations (https://www.youtube.com/watch?v=rWzla8NNhlw).
449 Given the feedback we received, we should have included subtitles for those hard of hearing. In
450 addition, we have decided to develop a video with Spanish subtitles, for obvious reasons. We
451 also believe that the consent form can be further improved. Inability to access results or
452 confusion about the creation of a 23andMe account should be covered in the consent. The
453 cardboard boxes that the kits arrive in are superfluous waste when utilizing almost 10,000 in less
454 than 3 months. They should be eliminated. Future sampling events will have to address the
455 incongruity between the diversity of the extant population and those who choose to participate
456 if we are to serve the health needs of all members of our communities.

457

458 **<u>Conclusion</u>**

459 A preliminary pilot population health study was initiated and the first phase was to collect saliva
460 and genotype from approximately 10,000 northern Nevadans. This was completed in a
461 catchment area surrounding Reno, NV where a single tertiary care facility treats ~70% of the
462 patient population. The participant population was skewed to female and was less racially diverse
463 than the population. The genetic test was administered free-of-charge and the patient

16

464 population was representative of the socio-economic diversity in northern Nevada. More specific

465 results for geographic, environmental, social and genetic patterns will be elucidated as the study

466 proceeds. The intent here was to highlight both the process and aggregate results of the

467 participants as compared to the population from which that sample was obtained.

468

469 Declarations

470 Ethics approval

471 This study was approved by the University of Nevada, Reno Institutional Review Board (956068-

472 8). Each participant provided written, informed consent.

473 Consent for publication

474 Not applicable.

475 Availability of data and materials

476 De-identified participant demographic data and SNP frequencies used for the cardiac risk score

477 calculations are available from the corresponding author on reasonable request.

478 Competing interests

479 The authors declare that they have no competing interests.

480 Funding

484 Authors' contributions

485 JG, CG and TS conceived of the study. MH, JG, JM, BL made substantial contributions to design.
486 JM, BL, HR, SR, RR made substantial contributions to data analysis. JG, JM, CR, CG, MC, TS made
487 substantial contributions to interpretation of data. JG, JM, CR, CG, MC, TS made substantial
488 contributions drafting the manuscript. All authors made substantial contributions analyzing
489 data and revising the manuscript.

490 Acknowledgements

17

References

1. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? BMJ. 2011;343:d4163–3.

2. United Health Foundation. America's Health Rankings 2016.

3. Cuckler G, Martin A, Whittle L, Heffler S, Sisko A, Lassman D, et al. Health spending by state of residence, 1991-2009. Medicare Medicaid Res Rev. 2011;1:E1–E31.

4. Washoe County Community Health Needs Assessment 2015-17.

5. Appointment Plus [Internet]. Available from: https://www.appointment-plus.com/

6. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med. 2016;375:2349–58.

7. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. Nature Research; 2016;48:1284–7.

8. Loh P-R, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. Nat Genet. Nature Research; 2016;48:811–6.

9. Deng JE, Sham PC, Li MX. SNPTracker: A Swift Tool for Comprehensive Tracking and Unifying dbSNP rs IDs and Genomic Coordinates of Massive Sequence Variants. G3: Genes| Genomes| Genetics. 2016.

10. Bureau UC. Census.gov. Available from: https://www.census.gov/

11. Hedt BL, Pagano M. Health indicators: Eliminating bias from convenience sampling estimators. Statistics in Medicine. John Wiley & Sons, Ltd; 2011;30:560–8.

12. Rosenthal R, Rosnow RL. Artifacts in Behavioral Research. Oxford University Press; 2009.

13. Junghans C, Jones M. Consent bias in research: how to avoid it. Heart. 2007.

14. Jordan S, Watkins A, Storey M, Allen SJ, Brooks CJ, Garaiova I, et al. Volunteer Bias in Recruitment, Retention, and Blood Sample Donation in a Randomised Controlled Trial Involving Mothers and Their Children at Six Months and Two Years: A Longitudinal Analysis. Hills RK, editor. PLoS ONE. Public Library of Science; 2013;8:e67912.

15. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. Nature Research; 2016;48:1031–6.

16. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, et al. Defining and Setting National Goals for Cardiovascular Health Promotion and Disease Reduction.

526   Circulation. American Heart Association, Inc; 2010;121:586–613.

527   17. Investigators A. The Atherosclerosis Risk in Communities (ARIC) study: design and
528   objectives. American Journal of Epidemiology. 1989;129:687–702.

529   18. Berglund G, Elmståhl S, Janzon L, Larsson SA. The Malmo Diet and Cancer Study. Design and
530   feasibility. Journal of Internal Medicine. Blackwell Publishing Ltd; 1993;233:45–51.

531   19. Baber U, Mehran R, Sartori S, Schoos MM, Sillesen H, Muntendam P, et al. Prevalence,
532   impact, and predictive value of detecting subclinical coronary and carotid atherosclerosis in
533   asymptomatic adults: the BioImage study. J. Am. Coll. Cardiol. 2015;65:1065–74.

534   20. Ridker PM, Chasman DI, Zee RYL, Parker A, Rose L, Cook NR, et al. Rationale, design, and
535   methodology of the Women's Genome Health Study: a genome-wide association study of more
536   than 25,000 initially healthy american women. Clin. Chem. 2008;54:249–55.

537

538   **Figure Legends**

539

540   Figure 1. Demographic map of the Reno/Sparks metropolitan area indicating per cent of

541   households in zip code blocks below the poverty level.

542

543   Figure 2. Distribution of cardiac risk scores among the participants in the study. The scores are

544   calculated as in Khera et al. [6] and is described in the text. Risk score means from other studies

545   are indicated [17-20].

546

Table 1. 2015 Census estimate racial and ethnicity data for Northern Nevada, Washoe County, Nevada and the U.S.

	United States	Nevada	Northern Nevada	Washoe County
White	77.10%	75.70%	86.70%	85.10%
Black	13.30%	9.30%	2.20%	2.60%
Native American	1.20%	1.60%	2.90%	2.20%
Asian	5.60%	8.50%	4.40%	6.00%
Pacific Islander	0.20%	0.80%	0.50%	0.70%
Other	2.60%	4.10%	3.30%	3.50%
Hispanic or Latino	17.60%	28.10%	22.10%	23.90%

U.S. Census Bureau, Population Estimates Program (PEP), Vintage 2015, Accessed 16 May 2017 <https://www.census.gov/quickfacts>.

Table 2. Self-reported ethnicity data for study participants and all Renown patients

	Study Participants		All Renown Participants	
	Count	% of Cohort	Count	% of Cohort
White	6515	88.86%	986645	73.98%
Black	91	1.24%	39190	2.94%
Native American	43	0.59%	15219	1.14%
Asian	155	2.11%	30786	2.31%
Pacific Islander	15	0.20%	2969	0.22%
Other	143	1.95%	44311	3.32%
Missing data	134	1.83%	78515	5.89%
Hispanic or Latino	236	3.22%	136086	10.20%

Table 3. Location of residence of the study participants.

	Study Participants		All Renown Patients	
	Count	% of Cohort	Count	% of Cohort
Reno/Sparks	6418	83.18%	703656	50.52%
All other Washoe County	104	1.35%	19319	1.39%
All other Northern Nevada	1052	13.63%	229877	16.50%
All other	142	1.84%	440005	31.59%

Table 4. 2015 American Community Survey age for Northern Nevada, Washoe County, Nevada and the U.S.

Age	United States	Nevada	Northern Nevada	Washoe County
Under 5 years	6.30%	6.40%	5.90%	6.20%
5 to 9 years	6.50%	6.80%	6.50%	6.30%
10 to 14 years	6.50%	6.50%	6.20%	6.40%
15 to 19 years	6.70%	6.30%	6.00%	6.30%
20 to 24 years	7.10%	6.70%	6.60%	7.40%
25 to 34 years	13.50%	14.30%	13.20%	14.10%
35 to 44 years	12.80%	13.70%	12.00%	12.50%
45 to 54 years	13.90%	13.60%	13.70%	13.70%
55 to 59 years	6.60%	6.30%	7.10%	6.80%
60 to 64 years	5.80%	5.80%	6.80%	6.40%
65 to 74 years	7.90%	8.40%	10.30%	8.70%
75 to 84 years	4.30%	3.90%	4.30%	3.70%
85 years and over	1.90%	1.30%	1.50%	1.50%

U.S. Census Bureau, 2011-2015 American Community Survey 5-Year Estimates. Accessed 16 May 2017 <http://factfinder2.census.gov>.

Table 5. Age data for study participants and all Renown Patients

|	Study Participants		All Renown Patients	
	Count	% of Cohort	Count	% of Cohort
<18 years	0	0.00%	143451	10.30%
18-30 years	1052	13.63%	206333	14.81%
31-40 years	1546	20.04%	163433	11.73%
41-50 years	1308	16.95%	172804	12.41%
51-60 years	1476	19.13%	181403	13.02%
61-70 years	1533	19.87%	205401	14.75%
71-80 years	698	9.05%	190203	13.66%
>80 years	103	1.33%	129829	9.32%
Missing Data	0	0.00%	0	0.00%

Table 6. American community survey (2010-4) study of income / poverty level for Northern Nevada, Washoe County, Nevada and the U.S.

	United States	Nevada	Northern NV	Washoe County
Median Household Income	$53,889	$51,847	$54,150	$52,870
Households Below Poverty Level	5.34%	4.97%	5.05%	5.38%
Individuals Age 25+ Below Poverty Level	12.01%	12.17%	11.27%	11.89%

Table 7. Sex of study participants and all Renown patients.

	Study Participants		All Renown Patients	
	Count	% of Cohort	Count	% of Cohort
Female	5140	66.59%	685155	49.18%
Male	2571	33.31%	706272	50.70%
Unknown	3	0.04%	250	0.02%
Missing data	5	0.06%	1420	0.10%

Figure 1.

Percent Households Below Poverty Level

Poverty Level Households
<5%
6-10%
11-20%
21-40%
41-60%
61-100%
R Renown Medical Facilities
R Renown Regional Medical Center
0 1 2 3 4 Miles

Sources: Esri, HERE, DeLorme, USGS, Intermap, INCREMENT P, NRCan, Esri Japan, METI, Esri China (Hong Kong), Esri Korea, Esri (Thailand), MapmyIndia, NGCC, © OpenStreetMap contributors, and the GIS User Community
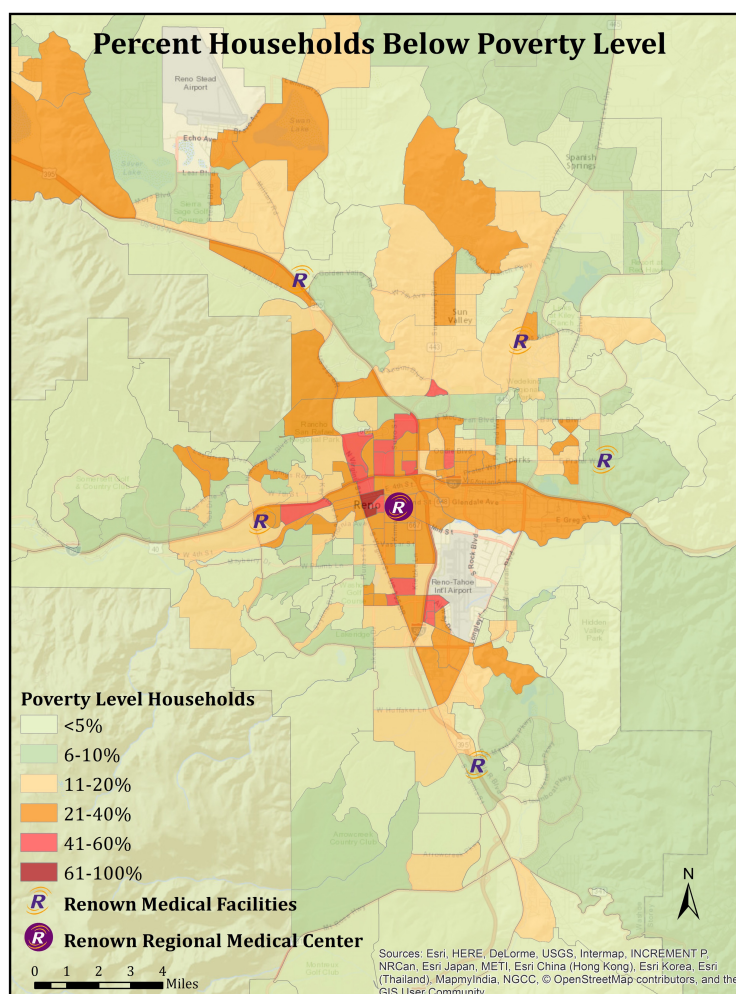
Figure 2.
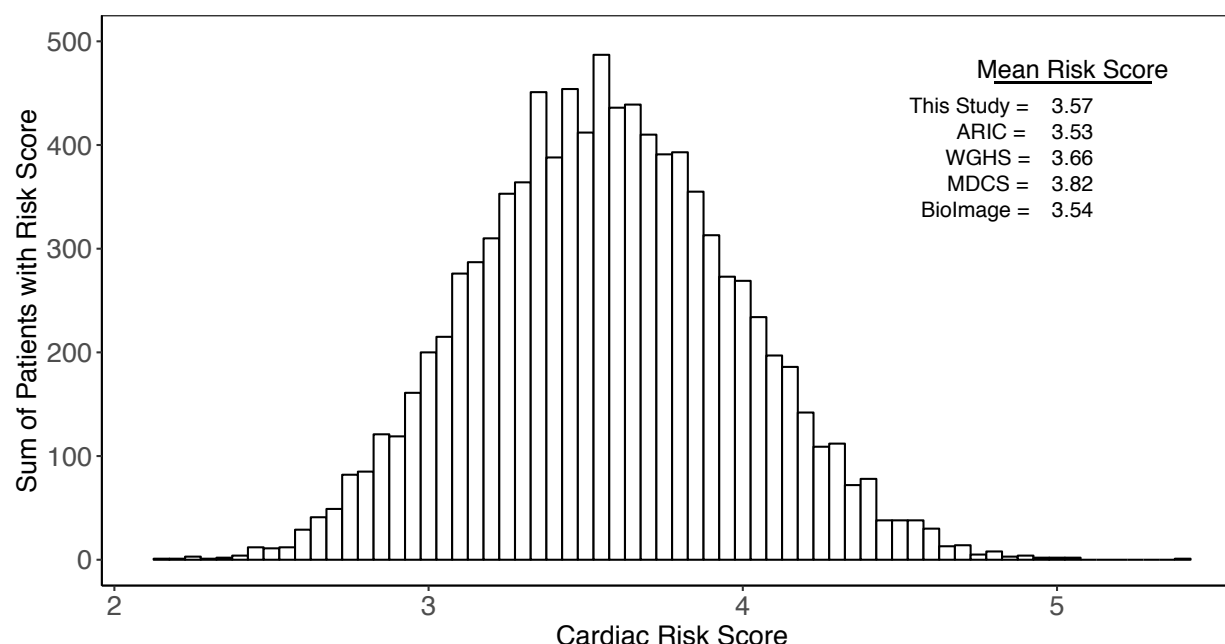
Supplemental Table 1. Location of residence of the study participants.

	Study Participants		All Renown Patients	
	Count	% of Cohort	Count	% of Cohort
Reno/Sparks	6418	83.18%	703656	50.52%
All other Washoe County	104	1.35%	19319	1.39%
All other Northern Nevada	1052	13.63%	229877	16.50%
All other	142	1.84%	440005	31.59%

Supplemental Table 2. 2015 American Community Survey age for Northern Nevada, Washoe County, Nevada and the U.S.

Age	United States	Nevada	Northern Nevada	Washoe County
Under 5 years	6.30%	6.40%	5.90%	6.20%
5 to 9 years	6.50%	6.80%	6.50%	6.30%
10 to 14 years	6.50%	6.50%	6.20%	6.40%
15 to 19 years	6.70%	6.30%	6.00%	6.30%
20 to 24 years	7.10%	6.70%	6.60%	7.40%
25 to 34 years	13.50%	14.30%	13.20%	14.10%
35 to 44 years	12.80%	13.70%	12.00%	12.50%
45 to 54 years	13.90%	13.60%	13.70%	13.70%
55 to 59 years	6.60%	6.30%	7.10%	6.80%

60 to 64 years	5.80%	5.80%	6.80%	6.40%
65 to 74 years	7.90%	8.40%	10.30%	8.70%
75 to 84 years	4.30%	3.90%	4.30%	3.70%
85 years and over	1.90%	1.30%	1.50%	1.50%

U.S. Census Bureau, 2011-2015 American Community Survey 5-Year Estimates. Accessed 16 May 2017 <http://factfinder2.census.gov>.

Supplemental Table 3. Age data for study participants and all Renown Patients

	Study Participants		All Renown Patients	
	Count	% of Cohort	Count	% of Cohort
<18 years	0	0.00%	143451	10.30%
18-30 years	1052	13.63%	206333	14.81%
31-40 years	1546	20.04%	163433	11.73%
41-50 years	1308	16.95%	172804	12.41%
51-60 years	1476	19.13%	181403	13.02%
61-70 years	1533	19.87%	205401	14.75%
71-80 years	698	9.05%	190203	13.66%
>80 years	103	1.33%	129829	9.32%
Missing Data	0	0.00%	0	0.00%

Supplemental Table 4. American community survey (2010-4) study of income / poverty level for Northern Nevada, Washoe County, Nevada and the U.S.

	United States	Nevada	Northern NV	Washoe County
Median Household Income	$53,889	$51,847	$54,150	$52,870
Households Below Poverty Level	5.34%	4.97%	5.05%	5.38%
Individuals Age 25+ Below Poverty Level	12.01%	12.17%	11.27%	11.89%

Supplemental Table 5. Sex of study participants and all Renown patients.

	Study Participants		All Renown Patients	
	Count	% of Cohort	Count	% of Cohort
Female	5140	66.59%	685155	49.18%
Male	2571	33.31%	706272	50.70%
Unknown	3	0.04%	250	0.02%
Missing data	5	0.06%	1420	0.10%

24