Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Genetic diversity in populations across Latin America: implications for population and medical genetic studies

Gillian M Belbin[1,2,3], Maria A Nieves-Colón[4,5], Eimear E Kenny[1,2,3], Andres Moreno-Estrada[4,8] and Christopher R Gignoux[6,7,8]

Hispanic/Latino (H/L) populations, although linked by culture and aspects of shared history, reflect the complexity of history and migration influencing the Americas. The original settlement by indigenous Americans, followed by postcolonial admixture from multiple continents, has yielded localized genetic patterns. In addition, numerous H/L populations appear to have signatures of pre-colonization and post-colonization bottlenecks, indicating that tens of millions of H/Ls may harbor signatures of founder effects today. Based on both population and medical genetic findings we highlight the extreme differentiation across the Americas, providing evidence for why H/Ls should not be considered a single population in modern human genetics. We highlight the need for additional sampling of understudied H/L groups, and ramifications of these findings for genomic medicine in one-tenth of the world's population.

Addresses
[1] The Center for Population Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[2] The Charles F Bronfman Center for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[3] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[4] National Laboratory of Genomics for Biodiversity (UGA-LANGEBIO), CINVESTAV, Irapuato, Guanajuato, Mexico
[5] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA
[6] Colorado Center for Personalized Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
[7] Department of Biostatistics and Informatics, School of Public Health, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA

Corresponding authors: Moreno-Estrada, Andres (andres.moreno@cinvestav.mx), Gignoux, Christopher R (chris.gignoux@ucdenver.edu)
[8] These authors contributed equally to this work.

Current Opinion in Genetics & Development 2018, **53**:98–104

This review comes from a themed issue on **Genetics of human origins**

Edited by **Lluis Quintana-Murci** and **Brenna Henn**

https://doi.org/10.1016/j.cogede.2018.07.006

Introduction

The intercontinental migrations of our recent evolutionary history have drastically changed the genetic makeup of human populations across the globe. Due to the confluence of highly diverged ancestry components, the admixture process in Latin America is considered one of the most pervasive forces having shaped diversity across two continents. Hispanic/Latino (H/L) populations are a heterogeneous group whose genetic roots trace their origin back to multiple continental and subcontinental lineages, with a modern population size over 700 million. Large-scale sampling efforts coupled with advancing genomic technologies have shed light on the admixture composition of H/L populations both at a continental scale [1–3], and regionally [4,5,6•].

A common finding across these studies and others is the strong evidence of a correlation between genetic patterns and geography, where the Native American fraction of admixed genomes shows greater affinities with local ethnic groups across the Americas. This is in part due to isolation of the ancestral groups soon after divergence and limited gene flow between them [7], but also because such ancestral groups were characterized by extremely small effective population sizes, creating a serial founder effect during the colonization of the Americas. Overall, this evolutionary pattern increased genetic differentiation and population diversification, resulting in a strongly substructured Native American population. Interestingly, such substructure is mirrored in the ancestry composition of admixed Latin Americans, supporting the importance of characterizing H/L populations at a finer scale beyond their continental ancestry proportions and questioning the simplified view of encompassing a continent-wide diversity in a single population group.

Population genetic history of Hispanic/Latinos

During Colonial times, not all Native American components contributed equally into the mixture of the emerging admixed population. In fact, some of the most isolated components may have not contributed at all, like some populations from desert or rainforest regions. Whereas others, like the descendants of big civilizations such as the Aztecs and the Mayans in Mesoamerica, or the Inca in Peru, have had a major impact in the composition of large present day populations across Latin America [8]. Likewise, only a fraction of all the possible sources from

Europe or Africa contributed to the mixture of H/L populations and, more importantly, they have played a differential role in shaping H/L genomes, with smaller contributing components having a potential stronger founder effect in the admixed population.

This mechanism is particularly relevant for disease mapping and medical genomic studies in H/L populations as the allele frequency of deleterious variants and missense mutations may have drastically changed from the patterns observed within its source population compared to that in H/L populations. Colonialism, through admixture and history record assimilation, has also favored the erasing of admixed individuals' origins, so much of the specific sources within the Americas, Europe, Africa and other contributing regions like Asia and the Middle East, remain largely uncharacterized. However, with the increasing availability of global genomic data and more sophisticated computational approaches to infer subcontinental ancestry components, these hidden ancestries are being brought to light by recent and ongoing studies aimed at filling the gaps in our understanding of H/L population genetic architecture, including detailed sampling of indigenous communities [9•]. Ongoing studies in other parts of the world therefore are likely to improve H/L genetic studies as well.

Deep structure in Hispanic/Latino groups enriches for founder populations

In population genetics, the 'founder effect' is the loss of genetic variation that occurs when a new population is established by a relatively small number of individuals. This process results in a randomly sampled genetic drift of variants relative to the parent population. In this scenario, potentially deleterious alleles can segregate at appreciable frequencies and these alleles can contribute substantially to the population specific disease burden of founder populations. Consequently, in genetic research much attention has been paid to populations that, due to geographical or cultural isolation, or patterns of endogamy, migration and diaspora, share signatures of a founder effect. In concert with the majority of genomic research, much of the focus has been on founder populations of European descent. Populations like the Icelandic, Sardinian, Finnish, and U.S. population isolates like the Amish and Hutterites, have been extraordinarily important for advancing Mendelian and complex trait genetics in European populations.

To date discovery efforts in non-European populations remain limited. This is especially true of the Hispanic/Latino (H/L) populations of the Americas. The term H/L encompasses a broad and highly heterogeneous group of populations. As mentioned, many of the contemporary H/L populations are the result of complex, recent demographic events including admixture, bottlenecking, and subsequent recent, rapid post-Colonial population growth. These demographic processes can result in founder effects that can create the conditions necessary for disease variants to drift to detectable frequencies. This suggests that, similar to European founder populations, researching H/L could be instrumental in driving discovery in both Mendelian and complex human disease. However, identifying these founder effects can be challenging due to the relative under representation of these population in genomic research databases, and that the underlying processes of gene flow do not necessarily correlate perfectly with the cultural, ethnic and regional labels that are typically used in research to categorize populations into discrete groups.

Some H/L groups have been studied via a classic founder population framework. For example the population of the Central Valley of Costa Rica (CVCR) has been studied as a genetic isolate population using haplotype based approaches to explore the genetic etiology of various complex disorders such as bipolar disorder, schizophrenia and asthma over the past two decades [10,11]. Recent population genetic surveys, however, have made discoveries in ostensibly cosmopolitan H/L populations that suggest founder effects may be more broadly ubiquitous across Latin America. For example in two studies [12,13], there was clear evidence of ancestry-specific bottlenecks, measured both in the indigenous and European components of ancestry in populations throughout the Caribbean and in the populations from the Americas in the 1000 Genomes Project. A recent study improved on these by systematically quantifying the relative magnitude of founder effects on different ancestral backgrounds in admixed H/L populations by estimating effective population size over time from the distribution of lengths of identical-by-descent (IBD) haplotypes stratified by different continental ancestral origin [14•]. Analyses of Colombian, Puerto Rican, Dominican, Ecuadorian, Guatemalan, Cuban, Honduran, Mexican and Nicaraguan populations derived from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) revealed variance in pre-admixture Ne, as well as a marked reduction in Ne around 12 generations ago, coincident with European contact with the Americas.

Another prominent H/L founder population are the Honduran Garifuna. Thought to be descendants of West African survivors of a shipwreck that crashed off the coast of St. Vincent in 1635, who subsequently admixed with the Island's Native Carib population before eventually being displaced to Central America [15], growing to over 1 million individuals today. Patterns of strong founder effect in Garifuna have been detected in multiple large independent population-scale studies inclusive of H/L groups. Genetic inference of the historical effective population size (Ne) of the Garifuna estimated using IBD supports this account; indicating that the population underwent a profound bottleneck ~12 generations ago,

with the minimum Ne inferred at 395 (95% CI: 352–466) [16•]. Similarly, analysis of mitochondrial data from Garifuna individuals revealed limited matrilineal diversity [17,18].

Across these studies then, it is clear that the number of H/L individuals with ancestry from founder populations is quite high, at least in the tens of millions (Individuals of Puerto Rican descent alone comprise over 9 million) to possibly hundreds of millions. In contrast to typical, smaller founder population studies in isolated regions, studying H/L groups provides an opportunity to investigate impacts of these founder effects on health outcomes. However, the number of founder population studies in H/L groups is small compared to European ancestry founder populations. Further, H/L's represent <4% of current large-scale genomic databases. The under representation of H/L diversity in such resources indicates a missed opportunity for genomic medicine discovery, potentially perpetuating health disparities.

Implications for medical genetics

Substructure at Mendelian variants

The complex demographic history of H/L populations has created the conditions necessary for rare variants to drift to higher frequencies in a highly geographically structured manner. This has important implications for the landscape of clinical variation across the Americas. To date myriad examples of clinical variants that are segregating in specific H/L populations, while being virtually absent elsewhere in the world have been reported. For example founder mutations in *BRCA1* and *BRCA2* have been reported in germline cases of breast cancer in Mexico [19], Chile [20] and Colombia [21]. Additionally, multiple founder mutations have been reported in individuals of Puerto Rican ancestry (e.g. [22,23]). Furthermore a Caribbean H/L specific founder variant in the *PSEN1* gene confers susceptibility to early onset Alzheimer's disease [24], notably a disease for which the population prevalence is higher in Caribbean H/L populations than in non-Hispanic White Americans. Another example of an H/L founder mutation is the *GHR*.pE180 variant underlying Laron Syndrome. This specific variant is only found in regions of Brazil, Chile and Ecuador, and has been observed to be segregating at appreciable frequency in the Loja region of Ecuador. Outside of H/L populations, this variant has only otherwise been observed among Sephardic Jews, and it has been postulated that the variant was brought to the Americas via a Sephardic founder effect [25]. With the growth of sequencing for medical diagnosis in patients with suspected genetic disorders in research and health systems, additional examples of founder mutations in H/L populations from across the Americas continue to be reported in the clinical literature.

Given the preponderance of founder effects in H/L groups, another model for Mendelian variant discovery that has emerged recently is the incidental discovery of founder mutations in large population-scale resources and biobanks. By combining population and statistical genetics approaches, researchers can first demonstrate genome-wide signatures of founder effects in sub-sets of the H/L population, and then discover founder disease mutations segregating there. For example, in the large diverse Bio*Me* Biobank recruited through the Mount Sinai Health System in New York City (NYC) (http://www.icahn.mssm.edu/research/ipm/programs/biome-biobank), over 30% of participants self-report as H/L. IBD sharing was elevated in H/L participants relative to other population groups, and a further analysis of IBD sharing in H/L sub-groups, showed the founder effects were pronounced in some groups and not others. Puerto Rican descent participants comprise one of the largest H/L founder populations in the NYC-based Bio*Me*. By linking to electronic health records, a subsequent IBD-haplotype based association study uncovered a Mendelian variant underlying the recessive disorder, Steel Syndrome. Population-level screening for this variant revealed that it segregates at a frequency of approximately 1–2% within Puerto Ricans, while being very rare or absent from elsewhere in the world. This suggests that this is a founder disease mutation, arisen to this frequency through processes of genetic drift, with important health implications for Puerto Ricans [26•]. Furthermore, this provides a new paradigm for incidental detection of some Mendelian disease variants segregating in H/L populations in large biobanks like the Bio*Me*, All of Us Research Program and the Million Veterans Program.

Substructure at complex trait variants

The observed degree of differentiation and isolation of indigenous groups, followed by heterogeneous migration from other continents, results in the expectation of high differentiation across modern H/L populations. Consistent with this, across much of the Americas, we see much higher Fst values than in other continents, even within countries. These numbers are amplified in many isolated indigenous groups, where smaller effective population sizes have resulted in increasing drift. For example, some population isolates within Mexico show up to 0.14 Fst values, which are higher than global inter-continental comparisons such as HapMap CEU versus ASN showing 0.11. Further, condensing population structure to a measure given by single allele frequencies does not capture the expected variance under admixture, as is almost ubiquitous in H/L groups.

It is natural then to assume that variants discovered via association studies will have heterogeneous patterns across Latin America. This can even affect patterns of replication in European-derived alleles, as the European component of H/L groups can be highly variable, and importantly,

experience ancestry-specific drift. In the investigation of complex traits, much of the seminal work on this has been performed in asthma, a disease with extreme differences in prevalence in Latin America, with Mexicans having lowest global prevalence and Puerto Ricans having the highest [27]. Consistent with this we observe that Mexicans and Puerto Ricans have heterogeneous patterns of association study replication [28], as well as unique signals for asthma and asthma-associated traits [29]. Similar patterns have been seen in other disease domains, such as the aforementioned identification of a collagen disorder founder mutation specific to Puerto Ricans, and a highly protective variant for breast cancer that is common in Mexicans [30], and a significant risk factor for Type 2 Diabetes [31,32]. These examples exhibit elevated allele frequencies in the indigenous ancestry component, and consistent with the demographic patterns observed above, are highly region-specific, demonstrating the importance of fine-scale study. This also implies that these variants would be unlikely to be observed in even the largest studies available of European-descent individuals.

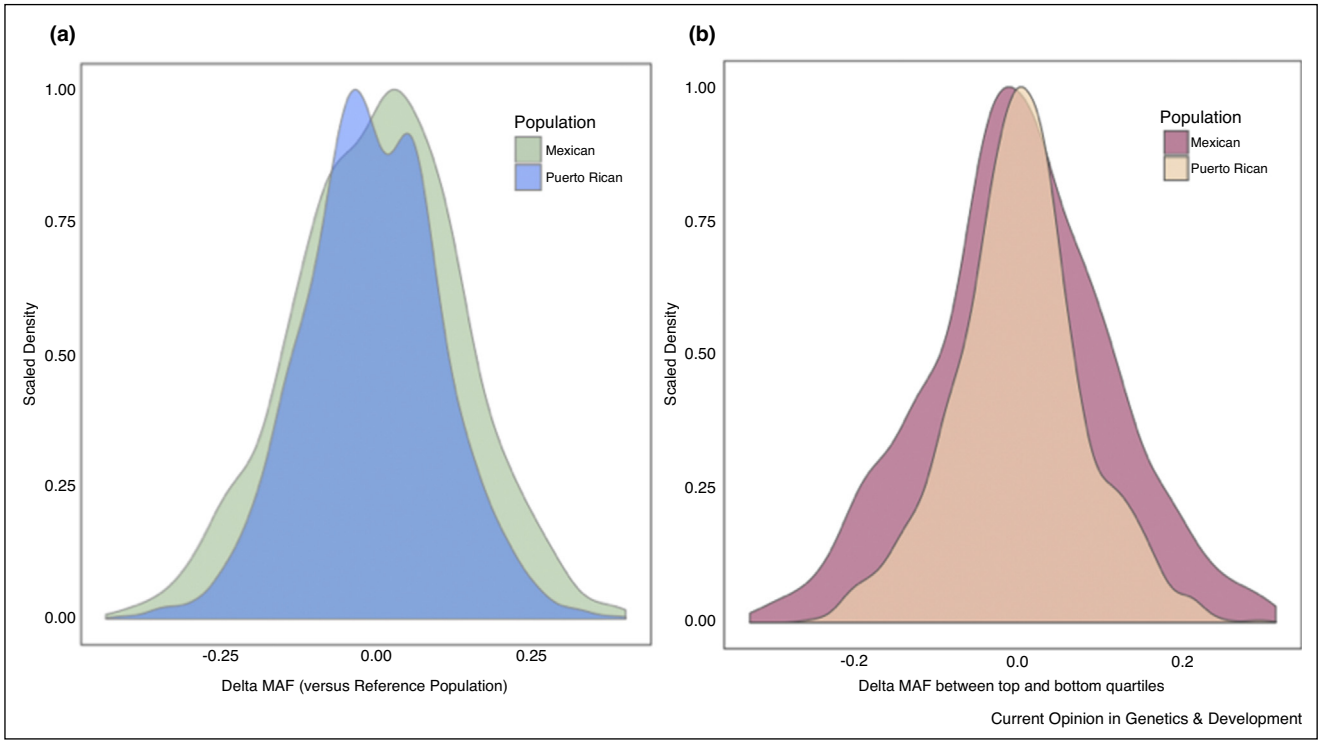
Extrapolating these to overall patterns of genome architecture for polygenic traits therefore leads to challenges in interpreting findings. It is well known that the field of human genetics suffers from an under-representation of non-European individuals in modern genomic databases [33,34••]. Previously, we have shown that this European-focused ascertainment results in biased polygenic scores in non-European populations [35••], and that modeling of fine-scale ancestry-specific patterns can improve personalized diagnostics in pulmonary function testing, a polygenic trait [4]. The problem is magnified in admixed populations. To test these results, we simulate admixed populations using binomial sampling. Briefly, we prespecify global ancestry patterns reflecting real-world population parameters, draw local ancestry states from those, then draw allele frequencies from the ancestral components, where parameterizing our draws using marginal beta distributions allows us to simulate *K*-way admixture. We begin with an ancestral causal allele, and derive the distribution of observed variants across the frequency spectrum for several populations, including Mexicans and Puerto Ricans. As observed in Figure 1, given the global Fst values observed across large populations in Latin America [36] we expect a large amount of drift affecting the distribution of allele frequencies. When compared to the ancestral reference population we observed that 35.8% of SNPs in the Puerto Rican simulation exhibited a difference in MAF that exceeded 10%, while for the Mexican simulation this was true for 45.5% of sites. These mean allele frequencies have a large effect and contribute significantly to the bias we described previously in polygenic scores [35••]. However, this is not the sole problem. If we look across the distribution, we note that the difference in allele frequencies between individuals in the top and bottom quartiles of the dominant ancestral component, even though they ostensibly are derived from the 'same population,' have a larger change in allele frequencies with 20.4% of sites exhibiting a difference of $\geq$10% when comparing individuals from the upper and lower quartiles for European ancestry in Puerto Ricans, and 36.0% of sites for Native American ancestry in the Mexican simulation. This then argues for a much more nuanced understanding of fine-scale population structure in the context of medical genetics in H/L groups. We note that this is a process shared by all groups that have received recent admixture, yet it is magnified by the multi-continental ancestry and local differentiation that underlies the genetic history of H/L populations.

Future research directions

Clearly, the observed heterogeneity across H/L groups necessitates further exploration and the need for linked population genetics and medical approaches. This is likely to be driven by three avenues. The first is the continuation of targeting specific understudied populations, such as African-descendant populations in South America (e.g. [37]). This is critical for understanding genetic patterns in unique, difficult-to-access populations, and for understanding localized differentiation that can affect medical discovery, consistent with observations seen in other parts of the world (e.g. [38]). In particular, the CANDELA project has proven highly successful as an international collaboration, leveraging population structure and anthropometric phenotypes to identify novel genetic associations [39•]. For groups with significant immigration into the United States, there have been opportunities for inclusion in large genetic studies, both those focusing on H/L groups [36] and large-scale multi-ethnic studies. Recent initiatives such as the NHGRI/NIMHD-sponsored Population Architecture using Genomics and Epidemiology Study (PAGE, www.pagestudy.org) and NHLBI Trans-Omics for Precision Medicine Program (TOPMed, www.nhlbiwgs.org), have incorporated tens of thousands of H/L individuals into multiple genomic investigations across a range of complex traits, while being cognizant of the inherent diversity present across multiple H/L groups.

A third opportunity for greater H/L inclusion lies in the growth of biobank repositories comprising electronical healthcare records and linked genetic data. Across the United States, we are observing a growing number with large H/L populations, including the Mount Sinai Bio*Me* Biobank. National projects such as the Million Veteran Program (http://www.research.va.gov/mvp) and The All Of Us Research Program (www.allofus.nih.gov) will include a large fraction of H/L participants, necessitating further investigation of the complexities of substructure and health disparities elucidated by a nuanced understanding of H/L genetics. However, as previously observed, understanding these patterns can contribute to novel insight into medical genomics and human biology. Further, similar efforts to build local capacity are being carried across Latin America,

Figure 1



(a) Difference in the allele frequency distribution between a simulated reference ancestral population (in this instance, European), and two daughter populations simulated using real-world parameters to represent Puerto Ricans and Mexicans. **(b)** Differences in the allele frequency distributions between individuals that fall within the upper versus lower quartiles of the dominant ancestral component per simulated population (namely European and Native American for Puerto Ricans and Mexicans, respectively).

including country-wide initiatives currently ongoing ranging from multi-institutional consortia in Mexico (www.mxbiobankproject.org) to Peru [6•], Brazil [40] and Chile (www.chilegenomico.cl).

Conclusions

The unique history of indigenous and mestizo populations present across Latin America proves to contribute a wealth of understanding to modern human genetics. Numerous high-impact discoveries have been made across a range of disease domains spanning rare Mendelian variants to unique aspects of complex trait biology. However to properly perform medical genetic studies, and empower future personalized medicine directions, studies in H/L populations require an understanding of fine-scale population structure and demography. As an example of this, we highlight the differentiation of H/L groups, the identification of founder effects in tens of millions of individuals, and the implications of these demographic patterns in medical genetics. Large-scale initiatives will provide exciting new opportunities for discovery, however we urge the field to embrace the inherent diversity present in H/L populations and not collapse them into a single category. Properly powered analyses, given modern tools and resources, can leverage this diversity for improved understanding of H/L populations, comprising roughly one-tenth of humanity, with ramifications for how we understand the genomes of all populations throughout the world.

Conflicts of interest statement

C.R.G. owns stock in 23andMe, Inc. C.R.G. and E.E.K. are founders and advisors to Encompass Bioscience, Inc. The remaining authors declare no conflict of interest.

Acknowledgements

We thank Noah Zaitlen for providing assistance with the initial simulation framework. This work was partially supported by the International Center for Genetic Engineering and Biotechnology (ICGEB) Grant CRP/MEX15-04_EC awarded to A.M.-E.; G.M.B. and E.E.K. were partially supported by NIH U01HG009080 and U01HG007417. C.R.G. and E.E.K. were partially supported by NIH R01HL104608. M.N.-C. was supported by the NSF Social, Behavioral & Economic Sciences Postdoctoral Research Fellowship (NSF award number 1711982).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- • of special interest
- •• of outstanding interest

1. Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM *et al.*: **Geographic patterns of**

genome admixture in Latin American Mestizos. *PLoS Genet* 2008, **4** e1000037.

2. Galanter JM, Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, Via M, Hidalgo-Miranda A, Contreras AV, Figueroa LU, Raska P *et al.*: **Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas**. *PLoS Genet* 2012, **8** e1002554.

3. Ruiz-Linares A, Adhikari K, Acuna-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, Fuentes M, Pizarro M, Everardo P, de Avila F *et al.*: **Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals**. *PLoS Genet* 2014, **10** e1004572.

4. Moreno-Estrada A, Gignoux CR, Fernandez-Lopez JC, Zakharia F, Sikora M, Contreras AV, Acuna-Alonzo V, Sandoval K, Eng C, Romero-Hidalgo S *et al.*: **Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits**. *Science* 2014, **344**:1280-1285.

5. Eyheramendy S, Martinez FI, Manevy F, Vial C, Repetto GM: **Genetic structure characterization of Chileans reflects historical immigration patterns**. *Nat Commun* 2015, **6**:6472.

6. • Harris DN, Song W, Shetty AC, Levano KS, Caceres O, Padilla C, Borda V, Tarazona D, Trujillo O, Sanchez C *et al.*: **Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire**. *Proc Natl Acad Sci U S A* 2018, **115**:E6526-E6535.

An example of a country-wide initiative to improve our understanding of local genetics, using a mix of whole genomes and genotype array data to address population genetics questions through before/after European contact, as well as implications for medical discovery.

7. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N *et al.*: **Reconstructing Native American population history**. *Nature* 2012, **488**:370-374.

8. Homburger JR, Moreno-Estrada A, Gignoux CR, Nelson D, Sanchez E, Ortiz-Tello P, Pons-Estel BA, Acevedo-Vasquez E, Miranda P, Langefeld CD *et al.*: **Genomic insights into the ancestry and demographic history of South America**. *PLoS Genet* 2015, **11** e1005602.

9. • Romero-Hidalgo S, Ochoa-Leyva A, Garciarrubio A, Acuna-Alonzo V, Antunez-Arguelles E, Balcazar-Quintero M, Barquera-Lozano R, Carnevale A, Cornejo-Granados F, Fernandez-Lopez JC *et al.*: **Demographic history and biologically relevant genetic variation of Native Mexicans inferred from whole-genome sequencing**. *Nat Commun* 2017, **8**:1005.

The first work reporting population-level high coverage genomes from indigenous Mexican communities, an important contribution to the broader understanding of genomics for individuals across Latin America.

10. Mathews CA, Reus VI, Bejarano J, Escamilla MA, Fournier E, Herrera LD, Lowe TL, McInnes LA, Molina J, Ophoff RA *et al.*: **Genetic studies of neuropsychiatric disorders in Costa Rica: a model for the use of isolated populations**. *Psychiatr Genet* 2004, **14**:13-23.

11. Morera B, Barrantes R: **Is the Central Valley of Costa Rica a genetic isolate?** *Rev Biol Trop* 2004, **52**:629-644.

12. Gravel S, Zakharia F, Moreno-Estrada A, Byrnes JK, Muzzio M, Rodriguez-Flores JL, Kenny EE, Gignoux CR, Maples BK, Guiblet W *et al.*: **Reconstructing Native American migrations from whole-genome and whole-exome data**. *PLoS Genet* 2013, **9** e1004023.

13. Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, Ortiz-Tello PA, Martinez RJ, Hedges DJ, Morris RW *et al.*: **Reconstructing the population genetic history of the Caribbean**. *PLoS Genet* 2013, **9** e1003925.

14. • Browning SR, Browning BL, Daviglus ML, Durazo-Arvizu RA, Schneiderman N, Kaplan RC, Laurie CC: **Ancestry-specific recent effective population size in the Americas**. *PLoS Genet* 2018, **14** e1007385.

A paper introducing a method that uses local ancestry deconvolution and IBD tract identification to quantify ancestry-specific bottlenecks, such as those that are commonly found in Latin American populations.

15. Crawford MH, Gonzalez NL, Schanfield MS, Dykes DD, Skradski K, Polesky HF: **The Black Caribs (Garifuna) of Livingston, Guatemala: genetic markers and admixture estimates**. *Hum Biol* 1981, **53**:87-103.

16. • Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, Vergara C, Torgerson DG, Pino-Yanes M, Shringarpure SS *et al.*: **A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome**. *Nat Commun* 2016, **7**:12522.

Quantifying admixture in primarily African-descent populations from across the Americas, including many Latin American groups. This paper quantifies the bottleneck mentioned in the Honduran Garifuna population, and provides ancestry estimates across a wide range of populations, along with high-coverage whole genomes.

17. Salas A, Richards M, Lareu MV, Sobrino B, Silva S, Matamoros M, Macaulay V, Carracedo A: **Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA**. *Am J Phys Anthropol* 2005, **128**:855-860.

18. Herrera-Paz EF, Matamoros M, Carracedo A: **The Garifuna (Black Carib) people of the Atlantic coasts of Honduras: population dynamics, structure, and phylogenetic relations inferred from genetic data, migration matrices, and isonymy**. *Am J Hum Biol* 2010, **22**:36-44.

19. Torres D, Rashid MU, Colombian Breast Cancer Study, Seidel-Renkert A, Weitzel JN, Briceno I, Hamann U: **Absence of the BRCA1 del (exons 9–12) mutation in breast/ovarian cancer families outside of Mexican Hispanics**. *Breast Cancer Res Treat* 2009, **117**:679-681.

20. Alvarez C, Tapia T, Perez-Moreno E, Gajardo-Meneses P, Ruiz C, Rios M, Missarelli C, Silva M, Cruz A, Matamala L *et al.*: **BRCA1 and BRCA2 founder mutations account for 78% of germline carriers among hereditary breast cancer families in Chile**. *Oncotarget* 2017, **8**:74233-74243.

21. Torres D, Rashid MU, Gil F, Umana A, Ramelli G, Robledo JF, Tawil M, Torregrosa L, Briceno I, Hamann U: **High proportion of BRCA1/2 founder mutations in Hispanic breast/ovarian cancer families from Colombia**. *Breast Cancer Res Treat* 2007, **103**:225-232.

22. Anikster Y, Huizing M, White J, Shevchenko YO, Fitzpatrick DL, Touchman JW, Compton JG, Bale SJ, Swank RT, Gahl WA *et al.*: **Mutation of a new gene causes a unique form of Hermansky-Pudlak syndrome in a genetic isolate of central Puerto Rico**. *Nat Genet* 2001, **28**:376-380.

23. Gonzaga-Jauregui C, Gamble CN, Yuan B, Penney S, Jhangiani S, Muzny DM, Gibbs RA, Lupski JR, Hecht JT: **Mutations in COL27A1 cause steel syndrome and suggest a founder mutation effect in the Puerto Rican population**. *Eur J Hum Genet* 2015, **23**:342-346.

24. Athan ES, Williamson J, Ciappa A, Santana V, Romas SN, Lee JH, Rondon H, Lantigua RA, Medrano M, Torres M *et al.*: **A founder mutation in presenilin 1 causing early-onset Alzheimer disease in unrelated Caribbean Hispanic families**. *JAMA* 2001, **286**:2257-2263.

25. Ostrer H: **The origin of the p.E180 growth hormone receptor gene mutation**. *Growth Horm IGF Res* 2016, **28**:51-52.

26. • Belbin GM, Odgis J, Sorokin EP, Yee MC, Kohli S, Glicksberg BS, Gignoux CR, Wojcik GL, Van Vleck T, Jeff JM *et al.*: **Genetic identification of a common collagen disease in Puerto Ricans via identity-by-descent mapping in a health system**. *Elife* 2017, **6**.

A work spanning population genetics and clinical informatics, uses the BioMe biobank data and the founder effect in Puerto Ricans to identify a large-effect allele for short stature via identity-by-descent mapping, and quantify its effects on health in carriers through phenome-wide approaches.

27. Lara M, Akinbami L, Flores G, Morgenstern H: **Heterogeneity of childhood asthma among Hispanic children: Puerto Rican children bear a disproportionate burden**. *Pediatrics* 2006, **117**:43-53.

28. Galanter JM, Torgerson D, Gignoux CR, Sen S, Roth LA, Via M, Aldrich MC, Eng C, Huntsman S, Rodriguez-Santana J *et al.*: **Cosmopolitan and ethnic-specific replication of genetic risk factors for asthma in 2 Latino populations**. *J Allergy Clin Immunol* 2011, **128** 37-43.e12.

29. Drake KA, Torgerson DG, Gignoux CR, Galanter JM, Roth LA, Huntsman S, Eng C, Oh SS, Yee SW, Lin L *et al.*: **A genome-wide association study of bronchodilator response in Latinos implicates rare variants**. *J Allergy Clin Immunol* 2014, **133**:370-378.

30. Fejerman L, Ahmadiyeh N, Hu D, Huntsman S, Beckman KB, Caswell JL, Tsung K, John EM, Torres-Mejia G, Carvajal-Carmona L *et al.*: **Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25**. *Nat Commun* 2014, **5**:5260.

31. Consortium STD, Estrada K, Aukrust I, Bjorkhaug L, Burtt NP, Mercader JM, Garcia-Ortiz H, Huerta-Chagoya A, Moreno-Macias H, Walford G *et al.*: **Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population**. *JAMA* 2014, **311**:2305-2314.

32. Consortium STD, Williams AL, Jacobs SB, Moreno-Macias H, Huerta-Chagoya A, Churchhouse C, Marquez-Luna C, Garcia-Ortiz H, Gomez-Vazquez MJ, Burtt NP *et al.*: **Sequence variants in SLC16A are a common risk factor for type 2 diabetes in Mexico**. *Nature* 2014, **506**:97-101.

33. Bustamante CD, Burchard EG, De la Vega FM: **Genomics for the world**. *Nature* 2011, **475**:163-165.

34. •• Popejoy AB, Fullerton SM: **Genomics is failing on diversity**. *Nature* 2016, **538**:161-164.
A landmark perspective that quantifies the degree to which the field, after a similar critique five years previous, has failed to close the gap in our understanding of diverse human population genetics, with ramifications for medical discovery.

35. •• Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE: **Human demographic history impacts genetic risk prediction across diverse populations**. *Am J Hum Genet* 2017, **100**:635-649.
A recent paper that demonstrates the degree to which European-focused ascertainment impairs our ability to apply polygenic risk scores across diverse populations, particularly large inherent biases due to genetic drift that affect the utility of polygenic scores as instruments for precision medicine.

36. Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, Nguyen EA, Drake KA, Huntsman S, Hu D *et al.*: **Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study**. *J Allergy Clin Immunol* 2014, **134**:295-305.

37. Conley AB, Rishishwar L, Norris ET, Valderrama-Aguirre A, Marino-Ramirez L, Medina-Rivas MA, Jordan IK: **A comparative analysis of genetic ancestry and admixture in the Colombian populations of Choco and Medellin**. *G3 (Bethesda)* 2017, **7**:3435-3447.

38. Nakatsuka N, Moorjani P, Rai N, Sarkar B, Tandon A, Patterson N, Bhavani GS, Girisha KM, Mustak MS, Srinivasan S *et al.*: **The promise of discovering population-specific disease-associated genes in South Asia**. *Nat Genet* 2017, **49**:1403-1407.

39. • Adhikari K, Fontanil T, Cal S, Mendoza-Revilla J, Fuentes-Guajardo M, Chacon-Duque JC, Al-Saadi F, Johansson JA, Quinto-Sanchez M, Acuna-Alonzo V *et al.*: **A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features**. *Nat Commun* 2016, **7**:10815.
An example of the discovery possibilities in the CANDELA project, a seminal, international collaboration to improve our understanding of anthropometric phenotypes across multiple populations in Latin America.

40. Kehdy FS, Gouveia MH, Machado M, Magalhaes WC, Horimoto AR, Horta BL, Moreira RG, Leal TP, Scliar MO, Soares-Souza GB *et al.*: **Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations**. *Proc Natl Acad Sci U S A* 2015, **112**:8696-8701.