# Clonal Hematopoiesis and the Risk for Rheumatoid Arthritis

Robert Corty, MD, PhD; 1
J. Brett Heimlich, MD, PhD; 2
Yash Pershad, BS; 3
Brian Sharber, MS; 3
Caitlyn Vlasschaert, MD, PhD; 4
Michelle Ormseth, MD, MSCI; 1
C. Michael Stein, MB, ChB; 5
Alexander Bick, MD, PhD; 3*

1 Department of Medicine, Division of Rheumatology, Vanderbilt University Medical Center, Nashville, TN, USA
2 Department of Medicine, Division of Cardiovascular Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
3 Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
4 Department of Medicine, Queen's University, Kingston, Ontario, Canada
5 Department of Medicine, Division of Clinical Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA
* Corresponding author:

Correspondence to:
Alexander G. Bick, MD, PhD
Division of Genetic Medicine, Department of Medicine
Vanderbilt University Medical Center
2200 Pierce Ave, Nashville, TN 37232
Phone: 615-322-4153
Email: alexander.bick@vumc.org

# Abstract

Objective: Rheumatoid arthritis (RA) is a systemic autoimmune disease with complex pathogenesis involving the innate and adaptive immune system. Clonal hematopoiesis of indeterminate potential (CHIP) is defined by clonal proliferation of one hematopoietic stem cell and is typically asymptomatic. Both are common among older adults. CHIP is associated with multiple autoimmune diseases, but has not been thoroughly evaluated for its relationship with RA.

Methods: We examined three large biobanks where CHIP status of participants has been determined from whole genome sequencing data. We ascertained cases of RA, seropositive RA (SPRA), and seronegative RA (SNRA) using established methods and used survival analysis to test whether CHIP status was predictive of incident disease. We combined the results of the three biobank studies using random effects meta-analysis. For validation, we performed deep, targeted sequencing of CHIP-causing genes in an established clinical cohort of 132 RA cases, 56 controls, and 544 external controls. We compared the rates of CHIP between cases and controls using logistic linear regression.

Results: In the UKBiobank and in meta-analysis, the presence of a large CHIP clone was associated with an increased risk for SPRA (HR = 2.57 with CI [1.46, 4.52] and p = 0.001) and RA (HR = 1.43 with CI [1.16, 1.75] and $p = 7 \times 10^{-4}$). Medium CHIP clones were associated with smaller increases in risk for SPRA and RA, and small CHIP clones carried no increased risk of any outcome. There were no associations detected between SNRA and CHIP of any size. In the clinical RA cohort, cases were more likely to have CHIP than controls after correcting for age, $age^2$, and sex (OR: 2.08, HR [1.09, 3.83], p = 0.02).

Conclusion: In a meta-analysis combing data from three large biobanks, large CHIP clones were associated with an increased risk for incident SPRA and, to a lesser extent, increased risk for RA. Validating this biobank-based finding, in a well-phenotyped clinical cohort, cases had higher rates of CHIP than age-matched controls. The mechanism by which CHIP drives the increased risk for SPRA is not known, but if it were discovered, could inform early intervention for patients with CHIP to prevent RA or personalized therapy for patients with RA based on CHIP status.

# Introduction

Rheumatoid arthritis (RA) is a systemic, autoimmune disease that primarily causes inflammatory arthritis.[1] The incidence of RA is approximately 100 cases per 100,000 person-years in women and half that in men with most cases arising in older adults.[2] Autoantibodies targeting the crystallizable fragment of the immunoglobulin G isotype, termed the "rheumatoid factor", and auto-antibodies targeting citrullinated portions of self-peptides are common in patients with RA, but their role in disease pathogenesis is not well understood.[6] Patients with RA and either of these types of auto-antibodies are considered to have "seropositive" RA (SPRA) and are at higher risk for bone erosion, interstitial lung disease, and treatment refractory disease than patients with "seronegative" RA (SNRA).

Although the cause of RA is not fully understood, many contributors are known. Cigarette smoking and mucosal infection with particular agents are risk factors for incident RA.[3] Common and rare genetic variants contribute through activating innate and adaptive immune pathways.[4,5] In aggregate, a complex interplay between genetic and environmental factors leads to innate and adaptive autoimmunity and a chronic, systemic autoimmune process that primarily affects joints, and can also affect the lungs, bursae, and sclerae, and carries an increased risk for cardiovascular disease.[7,8]

Clonal hematopoiesis of indeterminate potential (CHIP) is a blood condition defined by the presence of a leukemia-associated mutation in a hematopoietic stem cell leading to its clonal proliferation such that it produces 2% or more of the peripheral blood mononuclear cells.[9] CHIP is common among older adults, present in approximately 5% of people aged 60 and 15% of people aged 80, and typically asymptomatic.[10] However, people with CHIP have an elevated risk of myeloid neoplasm, cardiovascular disease, and all-cause mortality.[11,12] CHIP is associated with chronic inflammation, though the direction of causality between inflammation and CHIP is an area of active study.[13–15]

Given the complex interplay of innate and adaptive immunity that drives RA pathogenesis, CHIP may play a contributory role. The question of whether and how CHIP might contribute to RA pathogenesis has not been fully addressed. One study of 59 patients with RA found that 17% had CHIP.[24] Though this study did not contain matched controls, comparison to existing literature suggests that the rate of CHIP in each age group exceeded the population prevalence by approximately 10 percentage points.

We examined whether CHIP contributes to RA pathogenesis by testing whether CHIP is associated with risk for incident RA in three large biobanks and with RA cross-sectionally in an established clinical cohort.

# Materials and Methods

*Biobank Analyses*

Recognizing that CHIP and incident RA and CHIP are each individually uncommon, we studied participants in three large biobanks: The UKBiobank,[26] the NIH *AllOfUs* program,[27] the VUMC BioVU biobank[28] and combined their results in a meta-analysis. In each biobank, whole genome

sequencing (WGS) was performed by the centralized study team, and we identified patients with CHIP using established bioinformatic methods.[30]

*Inclusion Criteria*

To be eligible for study, biobank participants had to have:

1. whole genome sequencing data available,
2. at least one diagnosis code after blood sample collection date, and
3. no hematologic cancer prior to blood sample collection date.

Whole genome sequencing data was required because it is necessary to detect CHIP. At least one diagnosis code after the blood sample collection was required because the date of last medical contact was used as the right censoring date, so patients without any diagnosis codes after their blood sample would be interpreted to have no time at risk for developing RA. Patients with preceding hematologic cancer were excluded as it is an exclusion to being classified as having CHIP.

*Disease Ascertainment*

We ascertained patients with rheumatoid arthritis (RA), seropositive rheumatoid arthritis (SPRA), and seronegative rheumatoid arthritis (SNRA) using an established, validated approach with high specificity.[31,32] Specifically, we ascertained RA based on the following criteria:

1. **RA**: two or more diagnosis codes of RA, RF or CCP checked, and treated with an RA medication, as defined below.
2. **Sero-positive RA (SPRA)**: two or more diagnosis codes of SPRA, RF positive or CCP positive, and treated with at least one medication used to treat RA.
3. **Sero-negative RA (SNRA)**: two or more diagnosis codes of SNRA, RF negative or CCP negative (without any positive), and treated with at least one medication used to treat RA

ICD10 codes and SNOMED codes that indicate RA, SPRA, and SNRA are detailed in the supplementary materials. RF titer > 1:40 w and CCP titer > 1:20 was considered positive, so a patient with a laboratory value meeting either of these was considered seropositive. To be considered seronegative, a patient must have had at least one of RF and CCP checked, without any positive result. Medications that were considered indicative of RA in patients with RA diagnosis codes included hydroxychloroquine, sulfasalazine, methotrexate, leflunomide, infliximab, adalimumab, etanercept, certolizumab, and golimumab. Sensitivity analyses were conducted with relaxed definitions that required diagnosis codes plus (a) concordant laboratory values or medications (not necessarily both) and (b) diagnosis codes alone.

As this study used survival analysis, the date of first diagnosis with each disease was calculated. For RA, date of first diagnosis is the earliest date a relevant diagnosis code was assigned to the patient. For SPRA, the date of first diagnosis was the earlier of (1) date of first SPRA-specific diagnosis code or, for ascertainments that made use of seropositivity, (2) the date of first RA diagnosis code among patients who are proven by laboratory values to be seropositive. For SNRA, the date of first diagnosis was calculated similarly as to SPRA, except advancing it to the date of first RA diagnosis for patients proven to be seronegative by laboratory values.

*Survival Analysis*

In each biobank, Cox proportional hazards regression was used to estimate the hazard ratio for CHIP (and subsets thereof) on time to diagnosis with RA, SPRA, and SNRA, after correcting for sex,

race, age at biosample, age at biosample squared. Time at risk was considered to start on the date of biosample, and to end the date of first diagnosis with the disease being modeled. A given combination of CHIP-type and disease was considered to have adequate sample to model when there were at least 100 participants with the CHIP-type at baseline and at least 10 of them developed incident disease during the follow-up period. This analysis was conducted with the R programming language using package {survival}.[33] To assess the statistical significance of nominal *p* values in the context of the many hypotheses tested, *q* values were calculated using package {qvalue}.[34,35]

*Meta-analysis*
Random effects meta-analysis was conducted to combine the hazard ratios relating CHIP and RA across the three biobanks, using the restricted maximum likelihood estimator as implemented in package {metafor}.[36]

*Clinical Cohort study*
We followed up these biobank analyses with a re-examination of a Vanderbilt clinical cohort, a deeply-phenotyped cohort of patients with RA (n=132) and matched controls (n=56).[7] We performed deep, targeted sequencing of genes known to cause CHIP from frozen blood samples as has been described.[29] As the number of control samples in the RA clinical cohort was limited, we included additional controls (n=544) from the BioVU cohort, matched for age and sex, on which the same targeted sequencing assay had already been performed (unpublished data). Demographic features of cases, controls, and external controls are shown in Table 1. Cases with age < 55 or age > 75 were excluded due to low availability of both in-study and external controls. Using logistic regression to control for age, age$^2$, and sex, we estimated the odds ratio associated with CHIP positive status on the risk for RA. Due to the limited sample size, there was insufficient power to test the relationships between CHIP subsets such a specific driver genes or allele fractions and RA subsets such as seropositive and seronegative RA.

# Results

<u>UKB</u>
There were 404,735 participants in UKB eligible for study, based on eligibility criteria described above. Of these participants, 27,661 had CHIP as reported previously.[30]

After excluding participants with prevalent RA, the number of patients eligible for survival analysis for the specific ascertainment of RA, SPRA, and SNRA were: 402,033, 404,529, and 404,566 respectively and they accrued an aggregate time at risk of 7,447, 7,461, and 7,458 person-years. The number of incident diagnoses of RA, SPRA, and SNRA were: 522, 202, and 320, respectively. The rate of incident RA, SPRA, and SNRA was 7, 2, and 4 cases per 100 person-years, respectively.

CHIP with allele frequency (AF) 10% or greater and with AF 15% or greater were strongly associated an increased risk for SPRA (HR = 3.33 with CI [1.75, 6.32], $p = 2 \times 10^{-4}$ and $q = 7 \times 10^{-4}$ and HR = 2.57 with CI [1.46, 4.52], $p = 1 \times 10^{-3}$, $q = 3 \times 10^{-3}$ respectively). Additionally, CHIP with AF 5% or greater, 10% or greater, and 15% or greater was associated with an increased risk for RA (HR = 1.48 with CI [1.08, 2.04], $p = 0.01$, $q = 0.03$; HR = 1.97 with CI [1.33, 2.91], $p$ = 7 x 10-4, $q$ = 2 x 10-3; HR = 1.82 with CI [1.09, 3.04], $p = 0.02$, $q = 0.03$). Point estimates and confidence intervals for all tested CHIP-

RA associations are shown in Figure S2 and are not statistically significant after correcting for multiple hypothesis testing considerations.

AllOfUs

There were 152,852 participants in *AllOfUs* eligible for study, based on eligibility criteria described above. Of these participants 6,184 had CHIP as reported previously.[30]

After excluding participants with prevalent disease, the number of patients eligible for the most specific possible ascertainment of RA, SPRA, and SNRA were: 151,606 (specific), 151,706 (moderate), and 151,683 (moderate) and they accrued an aggregate follow-up time of 867, 877, and 876 person-years. The number of incident diagnoses were 191 (RA specific), 223 (SPRA moderate), and 266 (SNRA moderate). The rate of incident RA, SPRA, and SNRA was 22, 25, and 30 cases per 100 person-years, respectively.

No associations between CHIP of any size or gene specificity were statistically significant for any RA outcome(Figure S3).

BioVU

There were 111,581participants in BioVU eligible for study based on the eligibility criteria described above. Of these participants, 2,483 were found to have CHIP using established bioinformatic methods.[30]

After excluding participants with prevalent disease, the number of patients eligible for survival analysis for the most specific possible ascertainment of RA, SPRA, and SNRA were: 110,099 (specific), 110,721 (moderate), and 110,263 (moderate) who accrued a follow up time of 1,497, 1,511, and 1,494 person-years. The number of incident diagnoses were 666 (RA specific), 777 (SPRA moderate), and 1,516 (SNRA moderate). The rate of incident RA, SPRA, and SNRA was 44, 51, and 101 cases per 100 person-years, respectively.

No associations between CHIP of any size or gene specificity were statistically significant for any RA outcome (Figure S4).

Meta-Analysis

Meta analysis of the three biobank analyses found that CHIP with AF 10% or greater and CHIP with AF 15% or greater were strong risk factors for SPRA (HR = 2.57 with CI [1.46, 4.52] and p = 0.001; HR = 3.33 with CI [1.75, 6.32] and $p = 2 \times 10^{-4}$) as illustrated in Figure 1. Similarly, meta-analysis across the three biobanks found that CHIP with AF 10% or greater and CHIP with AF 15% or greater were risk factors for RA (HR = 1.32 with CI [1.06, 1.63] and p = 0.01; HR = 1.43 with CI [1.16, 1.75] and $p = 7 \times 10^{-4}$). Smaller CHIP clones with AF < 10% were not associated with risk for any disease and there were no associations between CHIP of any size with SNRA.

Clinical cohort

Of the 188 participants in the clinical cohort with biospecimens available, 132 were cases and 56 were controls. The cases and controls were of similar age distribution (54 +/- 10 vs 53 +/- 9) and sex distribution (69% female vs 61%, Table 2).

Given the limited number of controls in the original cohort, estimates of the rate of CHIP by age group among controls were imprecise. Therefore, we sought external controls from a previously

sequenced study to better estimate the rate of CHIP across the age range. After matching 10-to-1 on sex and age rounded to the nearest multiple of five, 544 external controls were identified (Table 2). External controls were older than in-study cases or controls (65 +/- 5 vs 54 +/- 10 and 53 +/- 9, $p < 0.001$ for both), due to the greater availability of controls near the upper end of the studied age range and less availability near the lower end as illustrated in Figure 2a.

Of the 132 RA cases, 19 carried CHIP (14%), while 98 of the 600 controls carried CHIP (16%) as illustrated in Figure 2b. Statistical modeling with the logistic linear regression model found that CHIP was associated with an odds ratio of 2.08 (CI 1.09 – 3.83) for RA after accounting for known effect of age on CHIP risk (Table 3, Figure 2c).

# Discussion

We studied over 650,000 research participants in three large biobanks: the UKBiobank, the NIH *AllOfUs* Research Program, and Vanderbilt's BioVU. We conducted survival analyses to test whether people who had CHIP at the time their blood was sampled were at higher risk of developing future RA or any specific subtype of RA as compared to people who did not have CHIP. We combined these single-cohort results using random effects meta-analysis.

A strong association between large CHIP clones (>10% or > 15%) and SPRA was noted in the UKB as well as an association between large CHIP clones and RA, though the hazard ratio for RA was lower. This result is consistent with the notion that CHIP confers an elevated risk for SPRA but not SNRA and therefore confers a lesser risk for RA, which includes the risk associated with SPRA diagnoses diluted by the lack of association with SNRA diagnoses. No statistically significant associations between CHIP and RA were identified in either the *AllOfUs* or BioVU cohorts individually. However, the *AllOfUs* cohort had approximately one third as many participants available for study as the UKBiobank, and much shorter follow-up time, so the absence of association may be due to lower power. Similarly, the BioVU cohort had approximately one third as many participants available for study as the UKBiobank, though the follow-up time was comparable to the UKBiobank. The meta-analysis results closely mirrored the results of the UKBiobank study, the most informative single-cohort study, and showed a strong association between large CHIP clones and SPRA.

In identifying CHIP from WGS, as we have done in each of these cohorts, the technical capability of the sequencing data is limited at the low end. WGS data in all three cohorts was performed at sequencing depth ~40 reads per genomic locus, which permits reliable detection of large CHIP clones but is susceptible to both false positive and false negative CHIP calls in the case of smaller clones (unpublished work).

The observed association between CHIP and SPRA without any observed association between CHIP and SNRA could reflect the biological reality that CHIP confers an increased risk for SPRA but not SNRA. Alternatively, it could be that, as SNRA is purely a clinical diagnosis without confirmatory objective data, it could be more often misdiagnosed and therefore patients who carry that diagnosis in biobank datasets are more likely to be misclassified.

We validated these biobank-based findings in the deeply phenotyped clinical cohort with deep, targeted genetic sequencing of the genes that cause more than 95% of all CHIP cases. We compared the rates of CHIP among patients with RA against both internal and external controls and

found that persons with RA are more likely to have CHIP than are healthy controls after controlling for the effect of age. The limited sample size in this cohort precluded the testing of the prevalence of specific types of CHIP, e.g. clones of particular sizes or driven by mutations in particular genes, or specific types of RA, e.g. SPRA or SNRA. This well-phenotyped cohort validated the overall findings from the biobank-based analyses.

Looking across biobank and cohort analyses, the strength of association between CHIP and RA was dependent on the fidelity with which both CHIP and RA could be ascertained. There was no association detected between difficult-to-detect small CHIP clones and any RA outcome, nor between CHIP clones of any size with difficult-to-ascertain SNRA. But a strong association was detected between large CHIP clones and SPRA, and a weaker association between large CHIP clones and RA, consistent with a dilution of the CHIP-SPRA signal into the CHIP-SNRA noise. Among the clinical cohort, where diagnoses were confirmed by rheumatology specialists using validated diagnostic criteria and sequencing was targeted and deep, the association between CHIP and RA was evident, even in the relatively small sample size.

Previous research has found CHIP to be associated with other autoimmune conditions. In biobank analyses, CHIP confers an increased risk for incident giant cell arteritis and gout.[17–20] In both cases, the increased risk for disease was caused mostly by CHIP clones driven by mutations in *TET2*. A reanalysis of a clinical trial in lupus found an elevated rate of CHIP among patients with lupus compared to their age-matched controls.[21] VEXAS is similar to CHIP in that it is driven by a clonal proliferation of cells carrying a somatic mutation, and 60% of patients with VEXAS have CHIP.[22] The role of somatic mutation in rheumatic disease is a rapidly-evolving field of research and this project adds RA to the growing list of CHIP-disease associations.[23]

In summary, RA has long been known to be a complex auto-immune disease that reflects pathophysiology in both the innate and adaptive immune systems. CHIP is both cause and consequence of chronic, mild hyperactivity of the innate immune system, so it is not entirely surprising that participants with CHIP clones of reliably detectable size carry an increased risk for SPRA, the more reliably diagnosed form of RA, and a diluted risk for RA in aggregate. The specific mechanisms by which CHIP drives this increased risk of SPRA are not known and merit further study as these insights might guide development of early interventions for patients with CHIP to avoid RA or precision interventions for patients with RA based on their CHIP status.

1. Gravallese, E. M. & Firestein, G. S. Rheumatoid arthritis - common origins, divergent mechanisms. *N. Engl. J. Med.* **388**, 529–542 (2023).

2. Conrad, N. *et al.* Incidence, prevalence, and co-occurrence of autoimmune disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK. *Lancet* **401**, 1878–1890 (2023).

3. Holers, V. M. *et al.* Rheumatoid arthritis and the mucosal origins hypothesis: protection turns to destruction. *Nat. Rev. Rheumatol.* **14**, 542–557 (2018).

4. Ishigaki, K. *et al.* Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat. Genet.* **54**, 1640–1651 (2022).

5. Kronzer, V. L., Sparks, J. A., Raychaudhuri, S. & Cerhan, J. R. Low-frequency and rare genetic variants associated with rheumatoid arthritis risk. *Nat. Rev. Rheumatol.* 1–11 (2024).

6. He, Y., Aoun, M., Xu, Z. & Holmdahl, R. Shift in perspective: autoimmunity protecting against rheumatoid arthritis. *Ann. Rheum. Dis.* (2024) doi:10.1136/ard-2023-225237.

7. Asanuma, Y. *et al.* Premature coronary-artery atherosclerosis in systemic lupus erythematosus. *N. Engl. J. Med.* **349**, 2407–2415 (2003).

8. Buch, M. H. *et al.* Current understanding and management of cardiovascular involvement in rheumatic immune-mediated inflammatory diseases. *Nat. Rev. Rheumatol.* (2024) doi:10.1038/s41584-024-01149-x.

9. Steensma, D. P. Clinical Implications of Clonal Hematopoiesis. *Mayo Clin. Proc.* **93**, 1122–1130 (2018).

10. Bick, A. G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).

11. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).

12. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N. Engl. J. Med.* **377**, 111–121 (2017).

13. Heimlich, J. B. *et al.* Multiomic Profiling of Human Clonal Hematopoiesis Reveals Genotype and Cell-Specific Inflammatory Pathway Activation. *Blood Adv* (2024) doi:10.1182/bloodadvances.2023011445.

14. Nathan, D. I., Dougherty, M., Bhatta, M., Mascarenhas, J. & Marcellino, B. K. Clonal hematopoiesis and inflammation: A review of mechanisms and clinical implications. *Crit. Rev. Oncol. Hematol.* **192**, 104187 (2023).

15. Belizaire, R., Wong, W. J., Robinette, M. L. & Ebert, B. L. Clonal haematopoiesis and dysregulation of the immune system. *Nat. Rev. Immunol.* **23**, 595–610 (2023).

16. Vlasschaert, C., Heimlich, J. B., Rauh, M. J., Natarajan, P. & Bick, A. G. Interleukin-6 Receptor Polymorphism Attenuates Clonal Hematopoiesis-Mediated Coronary Artery Disease Risk Among 451 180 Individuals in the UK Biobank. *Circulation* **147**, 358–360 (2023).

17. Robinette, M. L. *et al.* Association of Somatic TET2 Mutations With Giant Cell Arteritis. *Arthritis Rheumatol* (2023) doi:10.1002/art.42738.

18. Agrawal, M. *et al.* TET2-mutant clonal hematopoiesis and risk of gout. *Blood* **140**, 1094–1103 (2022).

19. Gutierrez-Rodrigues, F. *et al.* Clonal haematopoiesis across the age spectrum of vasculitis patients with Takayasu's arteritis, ANCA-associated vasculitis and giant cell arteritis. *Ann. Rheum. Dis.* (2023) doi:10.1136/ard-2023-224933.

20. Hecker, J. S. *et al.* CHIP and hips: clonal hematopoiesis is common in patients undergoing hip arthroplasty and is associated with autoimmune disease. *Blood* **138**, 1727–1732 (2021).

21. David, C. *et al.* Clonal haematopoiesis of indeterminate potential and cardiovascular events in systemic lupus erythematosus (HEMATOPLUS study). *Rheumatology* **61**, 4355–4363 (2022).

22. Gutierrez-Rodrigues, F. *et al.* Spectrum of clonal hematopoiesis in VEXAS syndrome. *Blood* **142**, 244–259 (2023).

23. Bucala, R. & Tsao, B. P. The emerging spectrum of somatic mutation in rheumatic disease: Clonal hematopoiesis connects aging with giant cell arteritis. *Arthritis Rheumatol* (2023) doi:10.1002/art.42745.

24. Savola, P. *et al.* Clonal hematopoiesis in patients with rheumatoid arthritis. *Blood Cancer J.* **8**, 69 (2018).

25. Hiitola, E. *et al.* Clonal Hematopoiesis Associated with Rheumatoid Arthritis. *Blood* **142**, 2698 (2023).

26. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).

27. All of Us Research Program Investigators *et al.* The "All of Us" Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).

28. Roden, D. M. *et al.* Development of a Large-Scale De-Identified DNA Biobank to Enable Personalized Medicine. *Clinical Pharmacology & Therapeutics* vol. 84 362–369 Preprint at https://doi.org/10.1038/clpt.2008.89 (2008).

29. Mack, T. *et al.* Cost-effective and scalable clonal hematopoiesis assay provides insight into clonal dynamics. *J. Mol. Diagn.* **26**, 563–573 (2024).

30. Vlasschaert, C., Mack, T., Heimlich, J. B. & Niroula, A. A practical approach to curate clonal hematopoiesis of indeterminate potential in human genetic datasets. *Blood* (2023).

31. Brooks, R. T. *et al.* The risk of lung cancer in rheumatoid arthritis and rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheumatol.* (2024) doi:10.1002/art.42961.

32. Chung, C. P., Rohan, P., Krishnaswami, S. & McPheeters, M. L. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. *Vaccine* **31 Suppl 10**, K41-61 (2013).

33. Therneau, T. & Grambsch, P. *Modeling Survival Data: Extending the Cox Model*. (Springer, New York, NY, 2010).

34. Storey, J. D. A Direct Approach to False Discovery Rates. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 479–498 (2002).

35. Storey, J. D., Bass, A. J., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. Preprint at http://github.com/jdstorey/qvalue (2022).

36. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* vol. 36 1–48 Preprint at https://doi.org/10.18637/jss.v036.i03 (2010).

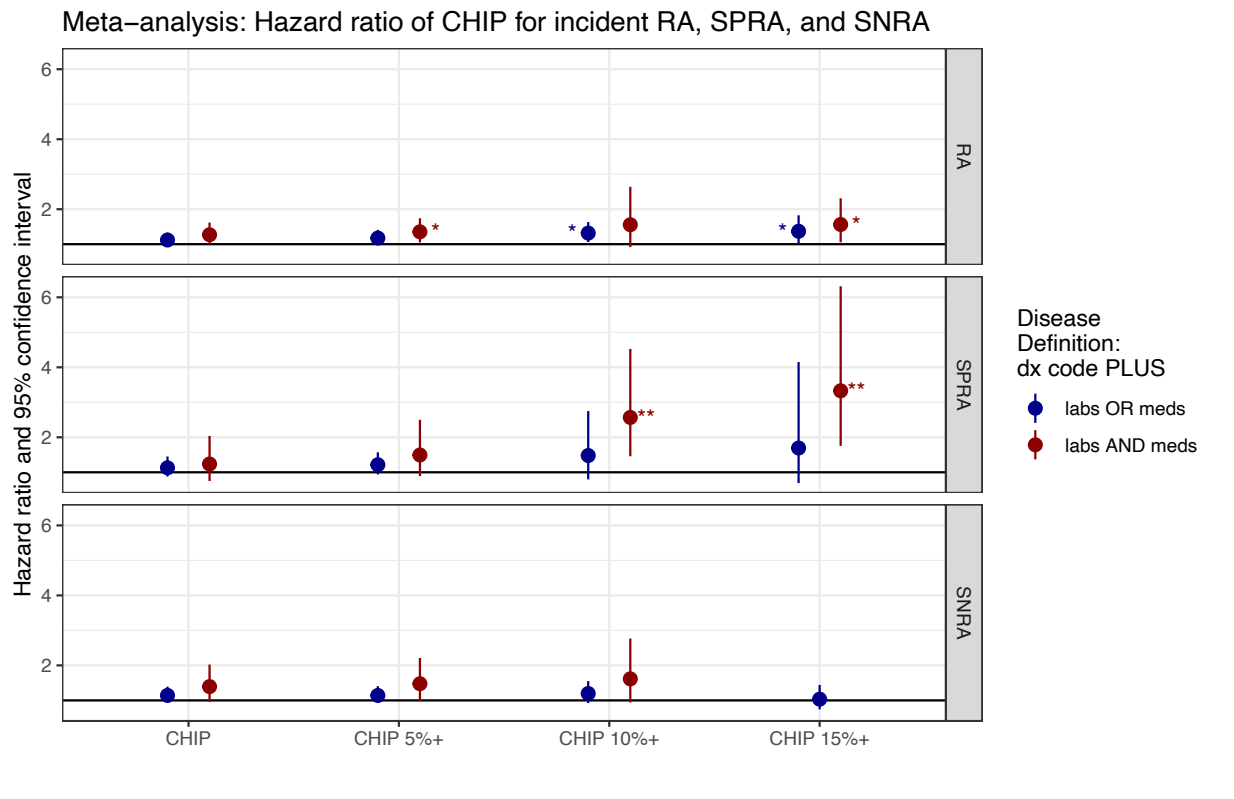


Figure 1: Meta-analysis across three cohorts of the effect of CHIP on the risk for incident RA, SPRA, and SNRA. Dots indicate point estimate of hazard ratio, vertical bars indicate 95% confidence interval for hazard ratio. There is a dose-dependent association between CHIP and incident RA, where CHIP with high VAF, e.g. 10%+ and 15%+, is associated with a large increase in the risk for SPRA and a statistically significant risk for all RA. * indicates nominal $p < 0.05$. ** indicates $p < 0.01$. There were not enough cases of participants with CHIP 15%+ developing SNRA to compute a reliable estimate of the hazard ratio.

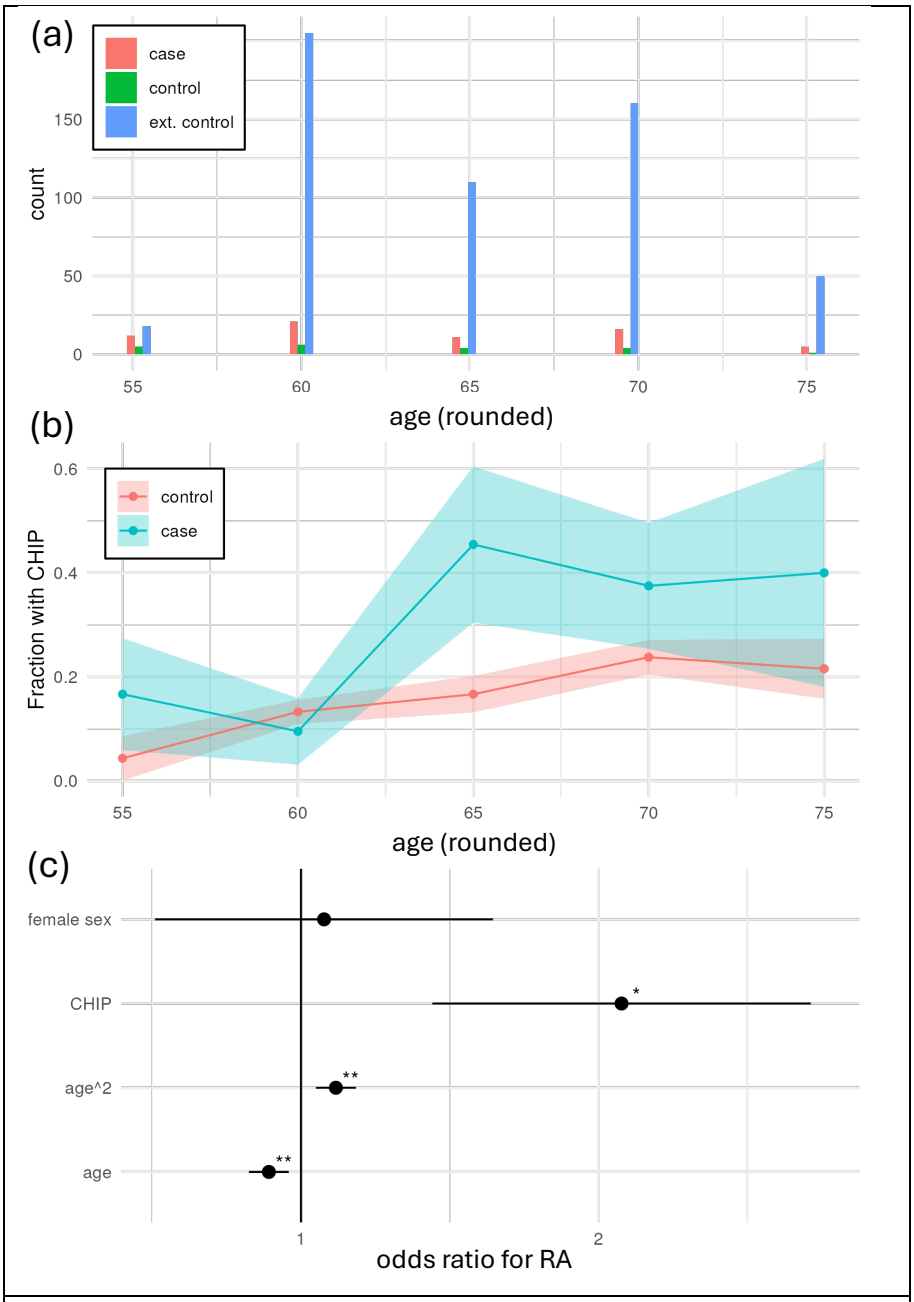


Figure 2: Results from clinical replication cohort. 2a: After combining the clinical replication cohort controls with matched external controls, there were more controls than cases in each age group. 2b: Across age groups, RA cases had modestly higher rate of CHIP than controls. 2c: CHIP is associated with a higher probability of RA in logistic linear regression analysis after correcting for age, age squared, and sex ($p = 0.02$).

| | | UKBiobank | NIH *AllOfUs* | BioVU |
|---|---|---|---|---|
| Patients included in study, n | | 404,735 | 152,852 | 111,581 |
| sex | Female, n (%) | 219,021 (54%) | 97,027 (64%) | 65,372 (59%) |
| race | White, n (%) | 383,210 (95%) | 87,570 (57%) | 75,381 (68%) |
| | Black or AA, n (%) | N/A | 27,235 (18%) | 27,554 (25%) |
| | Other, n (%) | 21,525 (5%) | 38,047 (25%) | 8,646 (8%) |
| Ever smoker | Yes, n (%) | 242,934 (60%) | 60,641 (40%) | 25,635 (23%) |
| Age at blood sample, mean (SD) | | 56 (8) | 53 (17) | 43 (21) |
| CHIP | Any CHIP, n (%) | 27,162 (6.7%) | 6,366 (4.2%) | 2,212 (2.0%) |
| | CHIP > 5%, n (%) | 21,326 (5.3%) | 6,339 (4.1%) | 2,212 (2.0%) |
| | CHIP > 10%, n (%) | 10,366 (2.6%) | 4,590 (3.0%) | 2,141 (1.9%) |
| | CHIP > 15%, n (%) | 6,187 (1.5%) | 2,178 (1.4%) | 1,620 (1.5%) |
| RA | Sensitive, n (%) | 4,862 (1.2%) | 4,077 (2.7%) | 4,201 (3.8%) |
| | Moderate, n (%) | 3,162 (0.8%) | 2,903 (1.9%) | 3,834 (3.4%) |
| | Specific, n (%) | 1,399 (0.3%) | 1,437 (0.9%) | 2,148 (1.2%) |
| SPRA | Sensitive, n (%) | 509 (0.1%) | 1,315 (0.9%) | 1,367 (1.2%) |
| | Moderate, n (%) | 1,115 (0.3%) | 1,369 (0.9%) | 1,637 (1.5%) |
| | Specific, n (%) | 599 (0.1%) | 635 (0.4%) | 870 (0.8%) |
| SNRA | Sensitive, n (%) | 397 (0.1%) | 702 (0.5%) | 2,402 (2.2%) |
| | Moderate, n (%) | 1,656 (0.4%) | 1,435 (0.9%) | 2,834 (2.5%) |
| | Specific, n (%) | 800 (0.2%) | 802 (0.5%) | 1,278 (1.1%) |

Table 1: Baseline and demographic information on three biobank populations. The number of patients ascertained to have moderate sensitivity SPRA and SNRA exceeds the number of patients with sensitive SPRA and SNRA due to the consideration of laboratory values permitting categorization of RA diagnosis codes into SPRA or SNRA, while the "sensitive" ascertainment uses only diagnosis codes.

| | case | control | external control |
|---|---|---|---|
| n | 132 | 56 | 544 |
| Age (mean (SD)) | 55 (10) | 53 (9) | 65 (5) |
| Female (n (%)) | 91 (69%) | 34 (61%) | 341 (63%) |
| RF positive (n (%)) | 89 (71%) | N/A | N/A |
| CHIP positive (n (%)) | 19 (14%) | 3 (5%) | 95 (17.5%) |
| CHIP 10%+ (n (%)) | 4 (3%) | 0 (0%) | 18 (3.3%) |

Table 2: demographic and baseline characteristics of clinical cohort cohort

| | Odds ratio | Confidence interval | *p*-value |
|---|---|---|---|
| Female sex | 1.08 | (0.62 – 1.91) | 0.79 |
| Age | 0.89 | (0.84 – 0.95) | < 0.001 |
| $Age^2$ | 1.12 | (1.05 – 1.19) | < 0.001 |
| CHIP | 2.08 | (1.09 – 3.83) | 0.02 |

Table 3: Odds ratios for RA in logistic regression analysis of the clinical cohort