RESEARCH ARTICLE

Elucidating the genetic architecture underlying IGF1 levels and its impact on genomic instability and cancer risk [version 1; peer review: 2 not approved]

Stasa Stankovic ¹, Felix R. Day ¹, Yajie Zhao ¹, Claudia Langenberg ¹, Nicholas J. Wareham¹, 23andMe Research Team, John R. B. Perry¹, Ken K. Ong ^{1,2}

¹MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

²Department of Paediatrics, University of Cambridge, Cambridge, UK

V1 First published: 01 Feb 2021, 6:20
<https://doi.org/10.12688/wellcomeopenres.16417.1>
 Latest published: 01 Feb 2021, 6:20
<https://doi.org/10.12688/wellcomeopenres.16417.1>

Abstract

Background: Insulin-like growth factor-1 (IGF1) has been implicated in mitogenic and anti-apoptotic mechanisms that promote susceptibility to cancer development and growth. Previous epidemiological studies have described phenotypic associations between higher circulating levels of IGF1 in adults with higher risks for breast, prostate, ovarian, colorectal, melanoma and lung cancers. However, such evidence is prone to confounding and reverse causality. Furthermore, it is unclear whether IGF1 promotes only the survival and proliferation of cancerous cells, or also the malignant transformation of healthy cells.

Methods: We perform a genome-wide association study in 428,525 white European ancestry individuals in the UK Biobank study (UKBB) and identify 831 independent genetic determinants of circulating IGF1 levels, double the number previously reported.

Results: Collectively these signals explain ~7.5% of the variance in circulating IGF1 levels in EPIC-Norfolk, with individuals in the highest 10% of genetic risk exhibiting ~1 SD higher levels than those in the lowest 10%. Using a Mendelian randomization approach, we demonstrate that genetically higher circulating IGF1 levels are associated with greater likelihood of mosaic loss of chromosome Y in leukocytes in men in UKBB (OR per +1 SD = 1.038 (95% CI: 1.010-1.067), $P=0.008$) and 23andMe, Inc. ($P=6.8 \times 10^{-5}$), a biomarker of genomic instability involved in early tumorigenesis. Genetically higher IGF1 is also associated with higher risks for colorectal (OR = 1.126 (1.048-1.210), $P=1.3 \times 10^{-3}$) and breast cancer (OR = 1.075 (1.048-1.103), $P=3.9 \times 10^{-8}$), with similar effects on estrogen positive (ER+) (OR = 1.069 (1.037-1.102), $P=2.3 \times 10^{-5}$) and estrogen negative (ER-) (OR = 1.074 (1.025-1.125), $P=3.9 \times 10^{-8}$) subtypes.

Open Peer Review

Approval Status

	1	2
version 1
01 Feb 2021	view	view
1. Joseph AMJL Janssen , Erasmus Medical Center, Molewaterplein, The Netherlands		
2. Rona Strawbridge , University Of Glasgow, Glasgow, UK		
Any reports and responses or comments on the article can be found at the end of the article.		

Conclusions: These findings give an insight into the genetic regulation of circulating IGF1 levels and support a causal role for IGF1 in early tumorigenesis and risks for breast and colorectal cancers.

Keywords

GWAS, loss of chromosome Y, IGF1, cancer, clonal mosaicism

Corresponding author: Stasa Stankovic (stasa.stankovic23@gmail.com)

Author roles: **Stankovic S:** Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **R. Day F:** Conceptualization, Data Curation, Software, Supervision; **Zhao Y:** Formal Analysis; **Langenberg C:** Resources; **J. Wareham N:** Resources; **R. B. Perry J:** Conceptualization, Formal Analysis, Funding Acquisition, Supervision, Writing – Review & Editing; **K. Ong K:** Conceptualization, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome (204845) and the Medical Research Council (MRC) (MC_UU_12015/1 and MC_UU_12015/2).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Stankovic S *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Stankovic S, R. Day F, Zhao Y *et al.* **Elucidating the genetic architecture underlying IGF1 levels and its impact on genomic instability and cancer risk [version 1; peer review: 2 not approved]** Wellcome Open Research 2021, 6:20 <https://doi.org/10.12688/wellcomeopenres.16417.1>

First published: 01 Feb 2021, 6:20 <https://doi.org/10.12688/wellcomeopenres.16417.1>

Introduction

Insulin-like growth factor-1 (IGF1) is a central component of an evolutionary conserved system that regulates the development and function of many tissues during embryonic development, growth and adulthood¹. While circulating IGF1 is produced primarily in the liver, IGF1 is expressed in almost every human tissue and plays a key role in mediating metabolic, endocrine and anabolic effects of growth hormone (GH)². GH stimulates the production of IGF1, which is used in clinical settings as an informative circulating marker for the diagnosis and monitoring of disease states of GH deficiency or excess. More widely, it has been posited as a potential marker with broader application for prediction of susceptibility of cancer and non-cancer aging traits³⁻⁵. Numerous epidemiological and clinical studies have described phenotypic associations between higher circulating levels of IGF1 in adults with higher risks for several cancers, including breast, prostate, ovarian, endometrial, colorectal and lung cancer, but also lower risk of ovarian cancer⁶⁻¹². However, phenotypic associations provide limited inference on causal effects, due to the potential effects of reverse causality and confounding. Studies of patients with acromegaly, a disease model of GH-IGF1 axis overactivity, indicate elevated risks for colorectal cancer and also for (small, low-risk) thyroid malignant nodules - most clinical guidelines recommend routine screening by colonoscopy in such patients although the evidence is far from equivocal¹³.

Findings from *in vivo* and *in vitro* experiments show that GH and IGF1 stimulate cell proliferation, differentiation and motility, and inhibit apoptosis¹⁴. The lifecycle of a healthy human cell is maintained through tight regulation of cell proliferation, senescence and apoptosis orchestrated by intracellular and extracellular signaling factors. Imbalance in this regulation that lowers apoptosis and stimulates higher proliferation rates leads towards more frequent mitosis, favouring carcinogenesis¹⁵. In searching for biomarkers that can be applied to predict and diagnose cancer, previous research has highlighted the potential roles of IGF1 in both cancer development and progression^{14,16}. Upon binding to the type 1 IGF receptor (IGF1-R), IGF1 stimulates a signal transduction cascade that results in higher proliferation rates and cell survival. This mitogenic activity is activated by RAS and AKT pathways that upregulate genes involved in cell proliferation, survival and angiogenesis¹⁷. Besides promoting mitogenesis, this cascade triggers downregulation of cell cycle suppressors, such as p27 and PTEN. The AKT pathway enables cancer cells to escape apoptosis mechanisms through the inhibition of pro-apoptotic proteins, such as BAD and FKHR, and activation of anti-apoptotic factors including NF-kappa B and MDM2¹⁸. Therefore, IGF1 may activate both mitogenic and anti-apoptotic pathways to promote cancer development and growth. However, it is debated whether its role is largely in promoting the survival, proliferation and metastasis of already abnormal cells, as opposed to driving genomic instability and malignant transformation of healthy cells.

Mendelian randomization (MR) provides a powerful genetic epidemiological approach to infer the causal effects of a candidate risk factor on an outcome, using genetic variants that

are robustly associated with that exposure as instrumental variables¹⁹. However, previous efforts to apply this approach to study the role of IGF1 in cancer have been inconclusive. The first genome-wide association studies (GWAS) of circulating IGF1 identified only 10 genome-wide significant signals in ~31,000 individuals of European ancestry^{20,21}. Two recent GWASs of circulating IGF1 in the UK Biobank study (UKBB) reported 265 independent signals in 194,174 women of European ancestry²² and 416 signals in 358,072 European descent individuals²³. The former study used these genetic signals in an MR framework to examine the role of IGF1 in breast cancer and found a positive effect of IGF1 on risk of estrogen receptor-positive (ER+) but not ER- breast cancer, while the latter study examined various cancer risks, and found a positive effect on risk of colorectal cancer but inconclusive or no evidence for prostate, breast, and other cancers.

Here, we conduct a GWAS of circulating IGF1 in UKBB using a more extensive approach to sample selection and signal selection than in previous studies²²⁻²⁵. We identify 831 independent signals for circulating IGF1 levels in white European ancestry adults, double the number of previously reported signals²³. We use this expanded genetic instrumental variable to make more conclusive inferences regarding the causal roles of IGF1 in colorectal, and breast cancers, and on mosaic loss of Y-chromosome in leukocytes, a biomarker of genomic instability.

Results

Discovery of IGF1 signals

After quality control procedures, genotype and phenotype data on circulating IGF1 levels were available in 428,525 white European individuals in UKBB. Genetic association testing, using a linear mixed model to control for relatedness and population structure, led to the discovery of 150,336 variants (MAF >0.1%) associated with IGF-1 levels at genome-wide significance ($p < 5 \times 10^{-8}$) (Figure 1). After distance based clumping and approximate conditional analysis, these resolved to 831 independent signals (*Extended data*, Supplementary Table 1)²⁶. We sought replication in an independent dataset (the EPIC-Norfolk study) of IGF-1 measurements in 7544 individuals²⁷. Although this study was substantially smaller than UKBB, we detected strong directional consistency of effects (in 601/791 autosomal signals; binomial sign test $P = 1 \times 10^{-50}$, including 118 at $P < 0.05$) (*Extended data*, Supplementary Table 2)²⁶. A GRS of 791 identified IGF1 signals explained ~7.5% of the variance in IGF1 levels in EPIC-Norfolk, and individuals in the highest 10% of the GRS had 0.96 SD [0.86-1.065] higher IGF 1 levels (5.5 nmol/L) compared to those in the lowest 10% of GRS. The GRS also explained 2.6% of the variance in IGF-binding protein- 3 (IGFBP3) levels in EPIC-Norfolk (N=3314), which is the major binding protein for IGF1.

Genetic correlations and biological pathways

The genetic architecture of IGF1 levels was highly correlated between men and women (genetic correlation: $r_g = 0.874$; $P < 1 \times 10^{-100}$). IGF1 was also more weakly genetically correlated with other complex health and behavioral traits, including positive correlations with adult height ($r_g = 0.086$; $P = 3.0 \times 10^{-4}$)

Figure 1. Manhattan plot of genome-wide association signals for circulating IGF1 levels in up to 428,525 white European individuals in UKBB. The association P value for each SNP (y axis) is plotted against its genomic position (x axis). The horizontal red line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Labelled gene names indicate signals tagged by deleterious variants.

and lung function (e.g. forced vital capacity: $r_g = 0.123$; $P = 2.9 \times 10^{-10}$), and negative correlations with obesity ($r_g = -0.116$; $P = 8.6 \times 10^{-7}$) and body fat, $r_g = -0.191$; $P = 4.5 \times 10^{-10}$) (*Extended data*, Supplementary Table 4)²⁶.

There was strong enrichment of GWAS signals in or near 60% of genes known to be involved in IGF1 core pathways, including IGF1 secretion, IGF1 serum balance and growth hormone secretion (*Extended data*, Supplementary Table 5)²⁶. We also performed two hypothesis-free approaches to explore pathways that contribute to circulating IGF1. First, we looked for nonsynonymous variants that are highly correlated (linkage disequilibrium $r^2 > 0.8$) with any lead IGF1 signal, and identified 220 such variants, tagging 165 unique genes (*Extended data*, Supplementary Table 3)²⁶. Predicted deleterious variants associated with IGF1 levels were identified in genes involved in the IGF-GH axis (*CREB3L2*, *GH1*, *GHRHR*, *IGFBP3*, *JAK2*, *MAPT*, *RASIP1*, *SH2B1*, *SSTR5*) and regulation of adrenal hormone secretion (*CRHR1*, *MC3R*, *PDE11A*, *POMC*), as well as cell growth, differentiation and death (*PTPN13*, *CDKN1A*, *ZC3HC1*). Secondly, we interrogated IGF1 associations at all variants genome-wide using MAGENTA and identified further pathways implicated in the regulation of IGF1 levels, including signaling by insulin, estrogen, and steroid hormone metabolism (*Extended data*, Supplementary Table 6)²⁶. These analyses also highlighted many pathways with less obvious links to

circulating IGF1 levels, including pathways responsible for neurodevelopment and immune response, more specifically viral response. Comparison between the pathways highlighted by MAGENTA and non-synonymous SNP analysis showed 79% overlap, indicating good consistency between these distinct approaches.

Examining the impact of IGF1 on genomic instability

We combined the identified 831 independent signals into a single genetic instrument in a MR framework in order to estimate the causal relationship between circulating IGF-1 levels and loss of chromosome Y (LOY) in leukocyte DNA in men. LOY represents a marker of genomic instability and processes involved in early tumorigenesis^{28,29}. In UKBB, each 1 SD (5.5 nmol/L) genetically predicted increase in IGF1 levels conferred a higher odds of LOY, measured using PAR-LOY (OR = 1.038; 95% CI: 1.010-1.067, $P = 0.008$ in inverse-variance weighted (IVW) model, with directionally consistent associations in sensitivity models and non-significant EGGER intercept $P = 0.56$) (*Extended data*, Supplementary Table 7)²⁶. Conversely, we found no evidence for a causal effect of LOY on IGF1 levels ($P = 0.09$). To confirm the validity of identified relationship between IGF1 and LOY, we performed a replication analysis in an independent dataset available for 653,019 male participants of European ancestry from the consumer genetics company 23andMe, Inc. Using a continuous measure

of Y-chromosome relative intensity as the outcome (mLRR-Y), with negative mLRR values representing more loss of chromosome Y, we confirmed the effect of higher genetically predicted IGF1 levels on LOY (Beta: -0.0003, 95% CI: -0.0004 - -0.0002, $P=6.8\times 10^{-5}$), with consistent findings across the sensitivity analysis (*Extended data*, Supplementary Table 7)²⁶.

Examining the impact of IGF1 on cancer risks

We then applied the same genetic instrument for circulating IGF1 levels to examine the potential causal influences of IGF1 on various cancer risks in published large-scale cancer consortium data and also UKBB data, which included smaller numbers of cases but covered a wider range of cancers.

Colorectal and thyroid cancers are more common in disease models of GH-IGF1 axis overactivity¹³ and therefore may be regarded as positive controls for this MR approach. Unfortunately, there were too few cases of thyroid cancer (N=375 in UKBB only) to provide meaningful MR estimates. However, we found highly consistent effects of IGF1 on higher risk for colorectal cancer across IVW and all sensitivity models (IVW

model: OR = 1.126; 1.048- 1.210, $P=1.3\times 10^{-3}$; N=5486 cases in UKBB only).

For breast cancer in women, in consortium data we identified a significant effect of higher IGF1 levels on higher risk of any breast cancer (OR per +1SD = 1.075; 95% CI: 1.048- 1.103, $P=3.9\times 10^{-8}$; N=122,977 cases), with similar effect sizes on ER+ (OR = 1.069; 1.037-1.102, $P=2.3\times 10^{-5}$) and ER- subtypes (OR = 1.074; 1.025- 1.125, $P=3.9\times 10^{-8}$) (Figure 2). Associations with ER+ breast cancer were robust across all sensitivity models (MR Egger, WM, PWM and MR Radial). The effect of IGF1 levels on ER- breast cancer was consistent in IVW models before and after MR Radial filtering, and was directionally consistent in other sensitivity models (*Extended data*, Supplementary Table 8)²⁶. In UKBB, a similar positive trend with any breast cancer was also identified (OR = 1.046; 0.998-1.097, $P=0.059$; N=13,666 cases) (Figure 3; *Extended data*, Supplementary Table 9)²⁶.

For prostate cancer in men, IGF1 levels had positive effects on cancer risk in IVW models both in consortium data (OR

Figure 2. Associations of genetic predisposition to higher IGF1 levels with site-specific cancers from large-scale consortiums. Odds ratios are per 5.5 nmol/l increase (95% CI) in IGF1. Results are obtained from the random-effects inverse-variance weighted method. OR: odds ratio; SD: standard deviation; CI: confidence interval.

Figure 3. Associations of genetic predisposition to higher IGF1 levels with site-specific cancers available in UKBB. Odds ratios are per 5.5 nmol/l increase (95% CI) in IGF1. Results are obtained from the random-effects inverse-variance weighted method post MR Radial analysis. OR: odds ratio; SD: standard deviation; CI: confidence interval.

per +1SD = 1.054; 1.017-1.093, $P=4.4 \times 10^{-03}$; $N=79,148$ cases) and in UKBB (OR = 1.068; 1.004-1.136, $P=3.7 \times 10^{-02}$; $N=7,872$ cases) before MR Radial filtering (Figure 2 and Figure 3). However, in both datasets this association showed moderate heterogeneity between variants ($I^2 = 58\%$ and 33% in consortia data and UKBB, respectively) and failed to reach significance in any sensitivity model (Extended data, Supplementary Table 9)²⁶.

Among other cancers that have been previously reported as phenotypically associated with circulating IGF1 levels⁶⁻¹², we found no genetic evidence linking circulating IGF1 levels to risks of endometrial, lung, or ovarian cancers.

We further explored causal influences of IGF1 on 17 other cancers in UKBB and consortia data as hypothesis generating. Genetically higher IGF1 levels conferred higher risk of multiple myeloma (OR = 1.261 per 1 SD; 1.026-1.550, $P=0.028$) with consistent findings after MR Radial filtering. Higher IGF1 was also associated with lower risks of biliary tract

(OR = 0.731; 0.559-0.956, $P=0.022$) and liver cancers (OR = 0.699; 0.522-0.937, $P=0.017$) both before and after MR radial filtering (Figure 3).

Multivariable MR models

IGF1 is a major driver of childhood growth and genetic associations with disease outcomes may therefore reflect lifetime IGF1 exposure. This is supported by our LD score regression analysis where genetic correlations with IGF1 levels were seen with anthropometric traits, including BMI ($r_g = -0.091$; $P=2.09 \times 10^{-5}$), body fat ($r_g = -0.191$; $P=4.50 \times 10^{-10}$) and height (height: females at age 10 and males at age 12, $r_g=0.157$; $P=3.34 \times 10^{-5}$). In addition, the MR analysis examining the effect of BMI on IGF1 levels showed significant causal association (OR 0.714, 95% CI 0.612-0.816, $P=0.016$), suggesting that BMI can act as a confounder of the effect of IGF1 levels on cancer risk.

Consequently, we performed multivariate MR analyses to distinguish causal effects of adult circulating IGF1 levels

from effects of adult height, as a summary indicator of childhood IGF1 exposure, and also adult BMI. In these multivariate models, genetically predicted IGF1 levels remained associated with higher risk of breast cancer (BMI-adjusted OR 1.068; 1.023-1.115, $P=0.003$; height-adjusted OR 1.069; 1.023-1.117, $P=0.003$), ER+ BrCa (BMI-adjusted OR 1.063; 1.014-1.114, $P=0.011$; height-adjusted OR 1.058; 1.008-1.109, $P=0.022$) and ER-BrCa (BMI-adjusted OR 1.062; 1.000-1.129, $P=0.05$; height-adjusted OR 1.081; 1.015-1.150, $P=0.015$). However, the genetic associations between IGF1 and risk of prostate cancer was attenuated on adjustment for BMI (OR 1.045; 0.988-1.105, $P=0.125$) (*Extended data, Supplementary Table 8*)²⁶.

Discussion

Here, we identified 831 independent genetic variants associated with circulating IGF1 levels, and used these variants to provide insights into the genetic regulation of IGF1 levels and the causal impact of IGF1 on early tumorigenesis and cancer risks. Our approach to GWAS sample inclusion and signal selection, using a combination of clumping and conditional analyses, led to twofold increase in the number of associated genetic determinants. The resulting stronger genetic instrument for IGF1 allowed us to detect evidence to support causal associations of IGF1 with LOY, a marker of genomic instability, and higher risks for colorectal and breast cancers, for which previous genetic studies had found inconclusive evidence²²⁻²⁵.

We showed that genetically predicted higher IGF1 levels were associated with higher risk of breast cancer, with similar effects on both ER+ and ER- subtypes, and consistent effects on any breast cancers in both BCAC and UKBB datasets. Recent MR studies using a smaller set of IGF1-associated variants found evidence for a role of IGF1 in ER+ but not ER- breast cancer^{22,23}. There is previous evidence that IGF1 may promote early risk factors for breast cancer, such as higher mammographic density³⁰, and also later stages of progression, such as proliferation and chemotherapy resistance of breast cancer cells³¹⁻³³. Our findings linking IGF1 and LOY support its role in early tumorigenesis. While LOY is measured in leukocytes and exclusively in men, recent studies have demonstrated that it represents ubiquitous DNA repair mechanisms and cell regulatory processes, with relevance to cancers in various tissues, and in both sexes, including higher risk of breast cancer in women^{28,29}.

Furthermore, previous genetic studies had suggested that any effect of IGF1 on breast cancer was restricted to ER+ subtypes^{22,23}, a view supported by experimental evidence for crosstalk in cells between the estrogens and IGF1 signaling pathways³⁴. Estrogen increases the expression of IGF receptors in breast cancer cells, thus increasing IGF1 mitogenic activity. By contrast, we showed that genetically predicted higher IGF1 levels conferred similar effects on both ER+ and ER- breast cancer subtypes, and suggests a mechanism for IGF1 independent of estrogen receptors to stimulate breast cancer formation.

Our finding for colorectal cancer is consistent with previous epidemiological studies^{25,35,36}, including a phenotypic analysis in UKBB which showed that higher prediagnostic levels of circulating IGF1 were associated with higher risks of proximal, distal and rectal colorectal cancers, which persisted controlling for other serologic factors (CRP, testosterone, SHBG, and HbA1c)²⁵. Colonocytes express IGF1-receptors; these are frequently overexpressed in neoplastic cells, and monoclonal antibody blockade of these receptors inhibits cell proliferation.

Previous phenotypic studies have reported the role of IGF axis in prostate cancer progression and mortality^{37,38}. We found a genetic association between IGF1 and risk of prostate cancer in both PRACTICAL and UKBB. However, no association was seen in any sensitivity analysis that controlled for the wide observed heterogeneity between effects of individual variants. Heterogeneity might arise because some variants are affected by confounding factors, such as BMI, or reflect differing effects of specific pathways in IGF1 regulation. For example, a previous MR analysis using a single genetic instrument in the *IGFBP3* region reported the opposite (i.e. protective) effect of IGF1 on prostate cancer progression and mortality²⁴. IGFBP3 binds 99% of circulating IGF1 and hence may protect tissues from effects of bioavailable IGF1 and may even have separate biological effects.

We acknowledge a number of limitations in our work. We identified preliminary evidence suggesting the role of IGF1 in multiple myeloma, hepatoma and biliary tract cancers. However, these outcomes were only available in UKBB, the same data source that we used to derive the genetic instrument for IGF1 levels. This is a potential source of bias and these findings require confirmation in future studies. Unfortunately, data on IGFBP3 levels were unavailable in UKBB, so we were unable to distinguish independent effects of IGF1 and its major binding protein. However, in the independent EPIC-Norfolk study, our IGF1 genetic score explained a much smaller proportion of variance in circulating IGFBP3 (~2.6%) than IGF1 (~7.5%) levels. Therefore, the disease effects indicated by our IGF1 genetic score are predominantly mediated by IGF1 alone. The MR approach assumes a linear relationship between the exposure and outcome. Hence, we could not test for threshold effects. However, previous phenotypic studies have found no evidence for such non-linear relationships, e.g. with risk of colorectal cancer²⁵.

Our findings may have potential implications for disease prevention and management. First, they support the further exploration of lifestyle approaches to avoid high IGF1 levels, for example, for example possibly by reducing total protein and dairy protein intakes^{39,40}. Secondly, our data underline existing clinical approaches for cancer screening in patients with disorders of GH-IGF1 axis overactivity⁴³, and may suggest extension of such protocols beyond colorectal cancer. Thirdly, our findings may encourage renewed interest in use of anti-IGF1 therapies for cancer outcomes^{41,42}.

In conclusion, our findings provide new insight into the genetic regulation of IGF1 levels, and demonstrate the likely

causal role of IGF1 in early tumorigenesis and susceptibility to breast and colorectal cancers, with preliminary evidence for other cancers.

Methods

Phenotype preparation in the UK Biobank

The genome-wide association study (GWAS) for IGF1 levels was performed in the UK Biobank study (UKBB)^{43,44}. Details of this study, including data collection and processing, are extensively described elsewhere^{45–47}. Briefly, UKBB involves ~500,000 adult participants with information on genotypes and phenotypes, including the measurements of 34 biomarkers. Informed consent was provided by all participants. Study approval was received from the National Research Ethics Service Committee North West–Haydock and all study procedures were performed according to the World Medical Association Declaration of Helsinki ethical principles for medical research. Here, we used genetic data from the UKBB ‘v3’ release⁴⁵, containing the full set of Haplotype Reference Consortium (HRC) and 1000 Genomes imputed variants⁴⁸.

Genetic data and IGF1 levels were available in up to 461,554 UKBB participants, with mean age 56.5 years (range: 37–73) and mean serum IGF1 concentration 21.4 nmol/l (range: 1.445–127.766 nmol/l). In addition to the quality control criteria applied by UKBB, we performed further procedures. In order to study a sample of relatively homogeneous ancestry, we restricted our GWAS analysis to UKBB participants of ‘white European’ ancestry, identified using K-means clustering for the first four principle components derived from genotype information. Individuals who clustered in this group but did not self-identify as white Europeans by the online questionnaire were excluded. Individuals with extreme levels of IGF1 (beyond 5 SD from the mean value) were also excluded. After these exclusions, data were available on up to 428,525 UKBB individuals, with a mean IGF1 concentration of 21.3 nmol/L (range: 1.445–42.449 nmol/l).

Genetic discovery analysis in the UK Biobank

The association between genetic variants and circulating IGF1 concentrations was examined using linear mixed models, implemented in **BOLT-LMM** v2.3.4, as this approach enables robust control for population stratification and cryptic relatedness⁴⁹. We performed three sets of association analysis, using male and female data, separately, adjusted for age, as well as a sex-combined dataset, adjusted for both age and sex. Statistically independent signals for IGF1 were identified using 1Mb distance-based clumping of all variants with P -value 5×10^{-8} , MAF $> 0.1\%$ and an imputation quality score > 0.5 . Lead signals were defined as independent signals with the lowest P -value, which were uncorrelated with other independent signals in each LD block ($r^2 < 0.05$). These were supplemented with additional independent signals (‘secondary signals’), which were identified using approximate conditional analysis implemented in genome-wide complex trait analysis (GCTA)⁵⁰. Conditional analysis considers LD between variants in a reference dataset, and selects additional signals independently associated with circulating IGF1 concentrations when conditioning on the effect of the lead signal at that locus. Secondary

signals met the following criteria: (1) $P < 5 \times 10^{-8}$ in both pre- and post-conditional analyses, (2) uncorrelated with another signal ($r^2 < 0.05$) and (3) its beta estimate changed by $< 20\%$ between pre- and post-conditional models⁵¹. The summary statistics and gene annotations for identified IGF-1 signals are shown in *Extended data*, Supplementary Table 1²⁶.

To assess the genetic correlation in circulating IGF1 concentrations between men and women, linkage disequilibrium score regression (LDSC) was applied to the summary statistics from the sex-stratified GWAS models⁵². The same approach was used to examine shared genetic aetiology between IGF1 and other complex health and behavioural traits using LD Hub, a centralized database and web interface which contains publicly available GWAS summary statistics for 245 traits of individuals with European ancestry⁵³.

Replication of GWAS identified signals was performed in an independent dataset from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk cohort, a population-based prospective study²⁷. Data on IGF1 levels were available for 7544 participants with genotype data only on autosomal variants. We examined consistency in the directions of association between individual signals and IGF1 levels, as well as their statistical significance. Where a particular signal was not present in EPIC-Norfolk, we searched the UKBB white European dataset for proxies (within 1 Mb and $r^2 > 0.5$) and chose the proxy with the highest r^2 value (*Extended data*, Supplementary Table 11)²⁶. Additionally, regression models were conducted on ventiles of the combined genetic score for higher IGF1 levels, and were controlled for assay batch, age, sex, BMI, height and ten genetic principal components as covariates. Allelic scoring was performed using the `--score` command in **PLINK** 1.9, weighted for effect estimates from the UKBB discovery analysis.

Identifying putatively functional genes

We applied the Ensembl **Variant Effect Predictor** (VEP)⁵⁴ to identify predicted deleterious variants correlated ($r^2 > 0.8$) with any IGF1 signal. The identified variants, including missense, stop-gained and splice site disrupting variants, were then classified in three categories, high, medium and low, based on the impact of a variant on the structure and function of a protein using VEP, **SIFT** and **PolyPhen**^{55,56}. ‘High impact’ variants were defined as stop gained, frameshift or splice site disrupting by VEP. ‘Medium impact’ variants were defined as: missense variants with moderate impact by VEP and deleterious by SIFT and at least possibly damaging by PolyPhen. ‘Low impact’ variants were defined as: missense variants with moderate impact by VEP and tolerated and/or benign by PolyPhen (*Extended data*, Supplementary Table 3)²⁶. LD was calculated from best guess genotypes for 1000 Genomes Phase 3 and HRC imputed variants using **PLINK** v1.9, based on 25,000 white British, unrelated participants from the UK Biobank. **UCSC LiftOver** was used to convert variant locations from b37 to b38.

Pathway analysis

We annotated GWAS signals located within 300 kb of genes known to be involved in the regulation of IGF-1, including three

specific pathways IGF1 secretion, growth hormone secretion and IGF1 serum balance, using KEGG, Reactome and manual curation. Secondly, we performed gene set enrichment analysis (GSEA) using STRING to identify pathways highlighted by the genes mapped by IGF1-associated non-synonymous variants (maximum distance 500kb).

Pathway associations were tested more broadly using [Meta-analysis Gene-set Enrichment of Variant Associations \(MAGENTA\)](#) software, which implements a GSEA-based approach⁵⁷. MAGENTA maps each gene in the genome to GWAS variants within upstream and downstream limits of 110 kb and 40 kb, respectively, and assigns each gene a score according to the lowest variant P-value for association with IGF1. Genes in the HLA region were excluded due to their strong and complex linkage disequilibrium patterns and high gene density. The gene score was then corrected for SNP density, LD-related properties and gene size in a regression model. In total, 3216 pathways from [Gene Ontology](#), [PANTHER](#), [KEGG](#) and [Ingenuity](#) were tested for enrichment of genes associated with IGF1 levels, using 10,000 permutations. Individual pathways were considered significant by a FDR<0.05 for either the 75th or 95th percentile for enrichment, and 90 pathways met this threshold (*Extended data*, Supplementary Table 6)²⁶.

Mendelian randomization: instrumental variable selection

MR analysis was applied to examine the likelihood of a causal effect of IGF1 on genomic instability and cancer risks. In this approach, genetic variants that are significantly associated with an exposure of interest are used as instrumental variables (IVs) to test the causality of that exposure on the outcome of interest. For a genetic variant to be a reliable instrument, the following assumptions should be met: (1) the genetic instrument is associated with the exposure of interest, (2) the genetic instrument should not be associated with any other competing risk factor that is a confounder, and (3) the genetic instrument should not be associated with the outcome, except via the causal pathway that includes the exposure of interest¹⁹.

We used the 831 IGF1 genome-wide significant signals as IVs. If a particular signal was not present in the outcome GWAS, we searched the UKBB white European dataset for proxies (within 1 Mb and $r^2 > 0.5$) and chose the variant with the highest r^2 value (*Extended data*, Table 11)²⁶. Genotypes at all variants were aligned to designate the IGF1-increasing alleles as the effect alleles.

MR: outcome data

Mosaic LOY in leucocyte DNA in men, an established marker of genomic instability, was used in a bidirectional MR analysis, thus examining the effect of IGF1 on LOY, and vice versa. This was performed using UKBB data, where the presence of LOY in individuals was identified using the pseudoautosomal region (PAR)-LOY method⁵⁸. This approach uses the diploid nature of the PAR to detect LOY based on the differences between maternal (X PAR) and paternal (Y PAR) allelic intensities at heterozygous sites. In the presence of LOY, Y PAR intensities are lower compared to X PAR intensities, and positive values

for PAR-LOY indicate more loss of chromosome Y. We used the genetic signals for LOY identified by Thompson *et al*, 2019 using the same PAR-LOY method²⁸ as genetic instruments to assess the causal inference of LOY on IGF1. The replication analysis was performed using the data generated from the customer base of 23andMe Inc., a consumer genetics company. This was done using a continuous measure of Y chromosome relative intensity as the outcome where LOY was quantified on the basis of median genotyping intensity over the non-PAR of the Y chromosome (mLRR-Y) using the protocol described previously²⁹. The data were available for 653,019 male participants of European ancestry. Unlike in PAR-LOY, negative values for mLRR-Y indicate more loss of chromosome Y.

According to the previously reported phenotypic evidence, we considered following primary cancer outcomes: colorectal, thyroid, breast, prostate, ovarian, melanoma and lung cancer¹⁴. There is a potential for bias in MR where IVs and outcomes are drawn from the sample. Therefore, where available we used independent cancer GWAS datasets from consortia without participant overlap with UKBB⁵⁹ and chose the largest publicly available GWAS dataset for each outcome. These include: the BCAC consortium for breast cancer (any cancer, and estrogen-receptor status specific cancers)⁶⁰, ECAC for endometrial cancer⁶¹, OCAC for ovarian cancer⁶², ILCCO for lung cancer⁶³ and PRACTICAL for prostate cancer⁶⁴. Information on the data sources is provided in *Extended data*, Supplementary Table 10²⁶.

We additionally assessed the causal inference of IGF-1 levels on cancer outcomes using data on 22 site-specific cancers available in UKBB for 367,586 unrelated white Europeans (*Extended data*, Supplementary Table 9)²⁶. Cancer data were available in UKBB from national cancer registries, electronic health records, hospital episode statistics data, death certification data, and self-reported information validated through an interview. In case of sex-specific cancers (breast, cervical, ovarian, prostate and uterine cancer) the data were available for individuals of the relevant sex only, including 198,838 women and 168,748 men).

MR models

Our primary MR model was the inverse-variance weighted (IVW) model, which offers the most statistical power⁶⁵. However, it does not correct for heterogeneity in outcome risk estimates between individual variants⁶⁶, which appeared to be significant for several cancer outcomes (*Extended data*, Supplementary Tables 8 and 9)²⁶. We therefore applied a number of sensitivity MR methods that better account for heterogeneity⁶⁷. MR Egger was used to identify and correct for unbalanced heterogeneity ('horizontal pleiotropy'), indicated by a significant Egger intercept ($P < 0.05$)⁶⁸. To correct for balanced heterogeneity, we used the weighted median (WM) and penalised weighted median (PWM) MR models⁶⁹. We also applied the MR Radial method which excludes variants from each model if they are recognized as outliers⁷⁰. Finally, we used Steiger filtering to assess for potential reverse causality (i.e. variants with stronger association with the outcome than with the exposure)⁷¹. This identified only one genetic variant (with a larger effect on endometrial cancer than on IGF1 levels), and excluding this

variant did not change the findings. I² statistics were used to assess heterogeneity in outcome risk estimates between individual SNPs.

Multivariable MR

Previous research on IGF1 showed its association with BMI and height: IGF1 is a major driver of childhood growth, thus adult data on height and BMI may represent a convenient indicator of childhood IGF1 exposure⁷². In order to confirm the effect of BMI on the levels of IGF1, we performed the primary MR analysis. BMI instruments were derived from the European sex-combined GIANT consortium dataset⁷³. After confirming the association, we investigated whether and to what extent BMI and height mediate the relationship between IGF1 and outcomes of interest using multivariable MR. To include BMI and height as covariates, we performed the lookup of IGF1 SNPs in BMI and height data available in the UKBB. The effects of each genetic instrument on multiple exposures are included as covariates, which after the adjustment allows the measurement of the independent, direct effect of IGF-1 levels on cancer outcomes not mediated by BMI and height⁷⁴.

Results presented in the main text are expressed per +1SD increase in IGF1 levels (equivalent to about 5.5 nmol/l in UKBB). Values in *Extended data*, Supplementary Tables 7–9²⁶ are raw data, per unit IGF1.

Data availability

Underlying data

Access to UK Biobank data is available to registered researchers worldwide from across academia and industry, without the need for collaboration, to perform health-related research that is in the public interest. The IGF-1 data from UKBB is available in the Discovery dataset: Data-Field 30772.

Similarly, EPIC-Norfolk (Replication study) data are available by application. To request data from EPIC-Norfolk, please [click here](#) to download a Data Request Form. To view a list of variable names collected by EPIC-Norfolk, then [click here](#) to access the Data Dictionary. To contact the EPIC-Norfolk team, please email epic-norfolk@mrc-epid.cam.ac.uk.

The data from 23andMe on loss of chromosome Y (LOY) used in our study are provided in the *Extended data*, Supplementary Table 12. Researchers wishing to access the full LOY dataset should follow the steps described on [23andMe](#) webpage, where they will be asked to submit a research proposal summary.

Extended data

Figshare: Elucidating the genetic architecture underlying IGF1 levels and its impact on genomic instability and cancer risk. <https://doi.org/10.6084/m9.figshare.13256333.v2>²⁶.

This project contains the following extended data:

- Supplementary Table 1. Genome-wide significant independent signals associated with levels of IGF1 and gene annotation.

- Supplementary Table 2. Replication Study in EpicNorfolk.
- Supplementary Table 3. Functional annotation of genome-wide significant signals and proxy variants.
- Supplementary Table 4. LD score regression analysis on the genetic correlations between IGF1 and other complex health and behavioural traits.
- Supplementary Table 5. IGF1 pathway analysis.
- Supplementary Table 6. Results of MAGENTA pathway analysis of genome-wide association data.
- Supplementary Table 7. Results of Mendelian Randomization analyses of genetically-mediated IGF1 levels on loss of chromosome Y.
- Supplementary Table 8. Results of Mendelian Randomization analyses of genetically-determined IGF1 levels (nmol/L) on cancer outcomes available in consortium studies.
- Supplementary Table 9. Results of Mendelian Randomization analyses of genetically-determined IGF1 levels (nmol/L) on cancer outcomes available in UKBB.
- Supplementary Table 10. Details of published genome-wide association studies used in Mendelian Randomization analyses of the effect of IGF1 levels on cancer outcomes.
- Supplementary Table 11. Weights for signals and proxies used in the genetic instrument for IGF1 levels.
- Supplementary Table 12. Weights for signals used as the genetic instrument for IGF1 levels, and their values in the 23andMe LOY dataset in the Mendelian Randomization framework.

Extended data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Acknowledgements

The authors acknowledge all participants and investigators involved in studies contributing to generation of summary statistics data for genetic associations with cancer, including the Breast Cancer Association Consortium, Endometrial Cancer Association Consortium, Ovarian Cancer Association Consortium, International Lung Cancer Consortium, Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome consortium and the UK Biobank. The research in the UK Biobank was conducted under application 9797. We also thank the research participants and employees of 23andMe, Inc. for making this work possible.

23andMe, Inc. Research Team

Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas

A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A. M. Northover, Steven J. Pitts,

J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic & Catherine H. Wilson

References

- Hakuno F, Takahashi SI: **40 YEARS OF IGF1: IGF1 IGF1 receptor signaling pathways.** *J Mol Endocrinol.* 2018; **61**(1): T69–T86.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Holly JMP, Perks CM: **Insulin-like growth factor physiology: what we have learned from human studies.** *Endocrinol Metab Clin North Am.* 2012; **41**(2): 249–63.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Marzullo P, Somma CD, Pratt KL, et al.: **Usefulness of Different Biochemical Markers of the Insulin-Like Growth Factor (IGF) Family in Diagnosing Growth Hormone Excess and Deficiency in Adults.** *J Clin Endocrinol Metab.* 2001; **86**(7): 3001–3008.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Brabant G: **Insulin-like growth factor-I: marker for diagnosis of acromegaly and monitoring the efficacy of treatment.** *European Journal of Endocrinology.* 2003; **148** Suppl 2: S15–S20.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Glynn N, Agha A: **Diagnosing Growth Hormone Deficiency in Adults.** *Int J Endocrinol.* 2012; 972617.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vigneri P, Tirrò E, Pennisi M, et al.: **The Insulin/IGF System in Colorectal Cancer Development and Resistance to Therapy.** *Front Oncol.* 2015; **5**: 530.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Allen NE, Key TJ, Appleby PN, et al.: **Serum insulin-like growth factor (IGF)-I and IGF-binding protein-3 concentrations and prostate cancer risk: results from the European Prospective Investigation into Cancer and Nutrition.** *Cancer Epidemiol Biomarkers Prev.* 2007; **16**(6): 1121–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kaaks R, Johnson T, Tikk K, et al.: **Insulin-like growth factor I and risk of breast cancer by age and hormone receptor status-A prospective study within the EPIC cohort.** *Int J Cancer.* 2014; **134**(11): 2683–90.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Yu H, Spitz MR, Mistry J, et al.: **Plasma levels of insulin-like growth factor-I and lung cancer risk: a case-control analysis.** *J Natl Cancer Inst.* 1999; **91**(2): 151–156.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lukanova A, Zeleniuch-Jacquotte A, Lundin E, et al.: **Prediagnostic levels of C-peptide, IGF-I, IGFBP -1, -2 and -3 and risk of endometrial cancer.** *Int J Cancer.* 2004; **108**(2): 262–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lukanova A, Lundin E, Toniolo P, et al.: **Circulating levels of insulin-like growth factor-I and risk of ovarian cancer.** *Int J Cancer.* 2002; **101**(6): 549–554.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Li Y, Li Y, Zhang J, et al.: **Circulating Insulin-Like Growth Factor-1 Level and Ovarian Cancer Risk.** *Cell Physiol Biochem.* 2016; **38**(2): 589–597.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Boguszewski C, Ayuk J: **MANAGEMENT OF ENDOCRINE DISEASE: Acromegaly and cancer: an old debate revisited.** *Eur J Endocrinol.* 2016; **175**(4): R147–R156.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Weroha J, Haluska P: **The insulin-like growth factor system in cancer.** *Endocrinol Metab Clin North Am.* 2012; **41**(2): 335–350, vi.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hanahan D, Weinberg R: **Hallmarks of Cancer: The Next Generation.** *Cell.* 2011; **144**(5): 646–674.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kasprzak A, Kwasniewski W, Adamek A, et al.: **Insulin-like growth factor (igf) axis in cancerogenesis.** *Mutat Res Rev Mutat Res.* 2017; **772**: 78–104.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rehman AG, Zwaalen M, Minder C, et al.: **Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis.** *Lancet.* 2004; **363**(9418): 1346–1353.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Samani AA, Yakar S, LeRoith D, et al.: **The role of the IGF system in cancer growth and metastasis: overview and recent insights.** *Endocr Rev.* 2007; **28**(1): 20–47.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smith GD, Ebrahim S: **'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?** *Int J Epidemiol.* 2003; **32**(1): 1–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kaplan RC, Petersen AK, Chen MH, et al.: **A genome-wide association study identifies novel loci associated with circulating IGF-I and IGFBP-3.** *Hum Mol Genet.* 2011; **20**(6): 1241–1251.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Teumer A, Qi Q, Nethander M, et al.: **Genomewide meta-analysis identifies loci associated with IGF-I and IGFBP-3 levels with impact on age-related traits.** *Aging cell.* 2016; **15**(5): 811–824.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Murphy N, Knuppel A, Papadimitriou N, et al.: **Insulin-like growth factor-1, insulin-like growth factor-binding protein-3, and breast cancer risk: observational and Mendelian randomization analyses with ~430 000 women.** *Ann Oncol.* 2020; **31**(5): 641–649.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Larsson S, Carter P, Vithayathil M, et al.: **Insulin-like growth factor-1 and site-specific cancers: A Mendelian randomization study.** *Cancer Med.* 2020; **9**(18): 6836–6842.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bonilla C, Lewis SJ, Rowlands MA, et al.: **Assessing the role of insulin-like growth factors and binding proteins in prostate cancer using Mendelian randomization: Genetic variants as instruments for circulating levels.** *Int J Cancer.* 2016; **139**(7): 1520–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Murphy N, Carreras-Torres R, Song M, et al.: **Circulating Levels of Insulin-like Growth Factor 1 and Insulin-like Growth Factor Binding Protein 3 Associate With Risk of Colorectal Cancer Based on Serologic and Mendelian Randomization Analyses.** *Gastroenterology.* 2020; **158**(5): 1300–1312.e20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Stankovic S, Day FR, Zhao Y, et al.: **Elucidating the genetic architecture underlying IGF1 levels and its impact on genomic instability and cancer risk.** *figshare.* Dataset. 2020.
<http://www.doi.org/10.6084/m9.figshare.13256333.v2>
- Day N, Oakes S, Luben R, et al.: **EPIC-Norfolk: study design and characteristics of the cohort.** *European Prospective Investigation of Cancer. Br J Cancer.* 1999; **80** Suppl 1: 95–103.
[PubMed Abstract](#)
- Thompson D, Genovese G, Perry J, et al.: **Genetic predisposition to mosaic Y chromosome loss in blood.** *Nature.* 2019; **575**(7784): 652–657.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wright DJ, Day FR, Kerrison ND, et al.: **Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility.** *Nat Genet.* 2017; **49**(5): 674–679.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lindström S, Thompson DJ, Paterson AD, et al.: **Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk.** *Nat Commun.* 2014; **5**: 5303.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Peyrat JP, Bonnetterre J, Hecquet B, et al.: **Plasma insulin-like growth factor-1 (IGF-1) concentrations in human breast cancer.** *Eur J Cancer.* 1993; **29A**(4): 492–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Endogenous Hormones and Breast Cancer Collaborative Group, Key TJ, Appleby PN, et al.: **Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies.** *Lancet Oncol.* 2010; **11**(6): 530–42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gu F, Schumacher FR, Canzian F, et al.: **Eighteen insulin-like growth factor pathway genes, circulating levels of IGF-I and its binding protein, and risk of prostate and breast cancer.** *Cancer Epidemiol Biomarkers Prev.* 2010; **19**(11): 2877–87.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dupont J, Le Roith D: **Insulin-like growth factor 1 and oestradiol promote cell proliferation of MCF-7 breast cancer cells: new insights into their synergistic effects.** *Mol Pathol.* 2001; **54**(3): 149–54.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rinaldi S, Cleveland R, Norat T, et al.: **Serum levels of IGF-I, IGFBP-3 and**

- colorectal cancer risk: results from the EPIC cohort, plus a meta-analysis of prospective studies.** *Int J Cancer.* 2010; **126**(7): 1702–15.
[PubMed Abstract](#) | [Publisher Full Text](#)
36. Verbanck M, Chen CY, Neale B, et al.: **Widespread pleiotropy confounds causal relationships between complex traits and diseases inferred from Mendelian randomization.** *BioRxiv.* 2017; 157552.
[Publisher Full Text](#)
 37. Chan JM, Stampfer MJ, Giovannucci E, et al.: **Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study.** *Science.* 1998; **279**(5350): 563–566.
[PubMed Abstract](#) | [Publisher Full Text](#)
 38. Travis R, Appleby PN, Martin RM, et al.: **A Meta-analysis of Individual Participant Data Reveals an Association Between Circulating Levels of IGF-I and Prostate Cancer Risk.** *Cancer Res.* 2016; **76**(8): 2288–2300.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Giovannucci E, Pollak M, Liu Y, et al.: **Nutritional Predictors of Insulin-like Growth Factor I and Their Relationships to Cancer in Men.** *Cancer Epidemiol Biomarkers Prev.* 2003; **12**(2): 84–89.
[PubMed Abstract](#)
 40. Bradbury K, Balkwill A, Tipper S, et al.: **The association of plasma IGF-I with dietary, lifestyle, anthropometric, and early life factors in postmenopausal women.** *Growth Horm IGF Res.* 2015; **25**(2): 90–95.
[PubMed Abstract](#) | [Publisher Full Text](#)
 41. Davis S, Eckhardt S, Diamond J, et al.: **A Phase I Dose-Escalation Study of Linsitinib (OSI-906), a Small-Molecule Dual Insulin-Like Growth Factor-1 Receptor/Insulin Receptor Kinase Inhibitor, in Combination with Irinotecan in Patients with Advanced Cancer.** *Oncologist.* 2018; **23**(12): 1409–e140.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 42. Kurzrock R, Patnaik A, Aisner J, et al.: **A Phase I Study of Weekly R1507, A Human Monoclonal Antibody Insulin-like Growth Factor-I Receptor Antagonist, in Patients with Advanced Solid Tumors.** *Clin Cancer Res.* 2010; **16**(8): 2458–2465.
[PubMed Abstract](#) | [Publisher Full Text](#)
 43. Sudlow G, Gallacher J, Allen N, et al.: **UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age.** *PLoS Med.* 2015; **12**(3): e1001779.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 44. Bycroft C, Freeman C, Petkova D, et al.: **Genome-wide genetic data on ~500,000 UK Biobank participants.** *bioRxiv.* 2017; 166298.
[Publisher Full Text](#)
 45. Bycroft C, Freeman C, Petkova D, et al.: **The UK Biobank resource with deep phenotyping and genomic data.** *Nature.* 2018; **562**(7726): 203–209.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 46. Elliott P, Peakman TC, UK Biobank: **The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine.** *Int J Epidemiol.* 2008; **37**(2): 234–244.
[PubMed Abstract](#) | [Publisher Full Text](#)
 47. UK Biobank Coordinating Centre: **UK Biobank: Protocol for a large-scale prospective epidemiological resource UK Biobank Coordinating Centre Stockport.** UKBB-PROT-09-06 (Main Phase) 06. 2007; 1–112.
[Reference Source](#)
 48. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al.: **A global reference for human genetic variation.** *Nature.* 2015; **526**(7571): 68–74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 49. Loh PR, Kichaev G, Gazal S, et al.: **Mixed-model association for biobank-scale datasets.** *Nat Genet.* 2018; **50**(7): 906–908.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 50. Yang J, Lee SH, Goddard ME, et al.: **GCTA: a tool for genome-wide complex trait analysis.** *Am J Hum Genet.* 2011; **88**(1): 76–82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 51. Yang J, Ferreira T, Morris A, et al.: **Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits.** *Nat Genet.* 2012; **44**(4): 369–375.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 52. Bulik-Sullivan B, Finucane H, Anttila V, et al.: **An atlas of genetic correlations across human diseases and traits.** *Nat Genet.* 2015; **47**(11): 1236–1241.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 53. Zheng J, Erzurumluoglu AM, Elsworth BL, et al.: **LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis.** *Bioinformatics.* 2017; **33**(2): 272–279.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 54. McLaren W, Gil L, Hunt SE, et al.: **The Ensembl Variant Effect Predictor.** *Genome Bio.* 2016; **17**(1): 122.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 55. Adzhubei I, Schmidt S, Peshkin L, et al.: **A method and server for predicting damaging missense mutations.** *Nat Methods.* 2010; **7**(4): 248–249.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 56. Vaser R, Adusumalli S, Leng S, et al.: **SIFT missense predictions for genomes.** *Nat Protoc.* 2015; **11**(1): 1–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
 57. Segrè A, DIAGRAM Consortium, MAGIC investigators, et al.: **Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits.** *PLoS Genet.* 2010; **6**(8): e1001058.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 58. Loftfield E, Zhou W, Graubard BI, et al.: **Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank.** *Sci Rep.* 2018; **8**(1): 12316.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 59. Pierce BL, Burgess S: **Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators.** *Am J Epidemiol.* 2013; **178**(7): 1177–1184.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 60. Michailidou K, Lindstrom S, Dennis J, et al.: **Association analysis identifies 65 new breast cancer risk loci.** *Nature.* 2017; **551**(7678): 92–94.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 61. Amos CI, Dennis J, Wang Z, et al.: **The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers.** *Cancer Epidemiol Biomarkers Prev.* 2017; **26**(1): 126–35.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 62. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al.: **Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer.** *Nat Genet.* 2017; **49**(5): 680–691.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 63. Wang Y, McKay JD, Rafnar T, et al.: **Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer.** *Nat Genet.* 2014; **46**(7): 736–741.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 64. Schumacher FR, Al Olama AA, Berndt SI, et al.: **Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci.** *Nat Genet.* 2018; **50**(7): 928–936.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 65. Slob EAW, Burgess S: **A comparison of robust Mendelian randomization methods using summary data.** *Genet Epidemiol.* 2020; **44**(4): 313–329.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 66. Bowden J, Fabiola Del Greco M, Minelli C, et al.: **A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization.** *Stat Med.* 2017; **36**(11): 1783–1802.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 67. Burgess S, Bowden J, Fall T, et al.: **Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants.** *Epidemiology.* 2017; **28**(1): 30–42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 68. Bowden J, Smith GD, Burgess S: **Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression.** *Int J Epidemiol.* 2015; **44**(2): 512–525.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 69. Bowden J, Davey Smith G, Haycock PC, et al.: **Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator.** *Genet Epidemiol.* 2016; **40**(4): 304–314.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 70. Bowden J, Spiller W, Del Greco M, et al.: **Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression.** *Int J Epidemiol.* 2018; **47**(6): 2100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 71. Hemani G, Tilling K, Davey Smith G: **Orienting the causal relationship between imprecisely measured traits using GWAS summary data.** *PLoS Genet.* 2017; **13**(11): e1007081.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 72. Larsson SC, Michaëlsson K, Burgess S: **IGF-1 and cardiometabolic diseases: a Mendelian randomisation study.** *Diabetologia.* 2020; **63**(9): 1775–1782.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 73. Locke AE, Kahali B, Berndt SI, et al.: **Genetic studies of body mass index yield new insights for obesity biology.** *Nature.* 2015; **518**(7538): 197–206.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 74. Sanderson E, Smith GD, Windmeijer F, et al.: **An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings.** *Int J Epidemiol.* 2019; **48**(3): 713–727.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 16 May 2022

<https://doi.org/10.21956/wellcomeopenres.18064.r50301>

© 2022 Strawbridge R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rona Strawbridge

Institute of Health and Wellbeing, University Of Glasgow, Glasgow, UK

The authors report a GWAS of IGF1 levels in the UK BioBank study, accompanied by replication in an independent dataset, genetic correlation analyses, risk scoring and mendelian randomisation. Whilst interesting, I have major concerns about the methodology used, the lack of biological consideration applied and therefore any conclusions drawn. Specifically, the use of the same dataset/individuals for GWAS discovery and MR analyses are inappropriate. The authors should take care when preparing a manuscript where the results are presented before the methods, that adequate information about the methods is provided so that the results can be interpreted.

Major comments:

1. The authors refer to two previous GWAS studies of the same phenotype in the same dataset. Aside from defining loci in a different way compared to those studies, how does this study differ from those studies and/or add to the literature?
2. There is no mention of descriptive statistics of the datasets used here, therefore it is not possible to tell if the datasets are even loosely comparable in terms of demographics and IGF1 methods/levels.
3. The "discovery of IGF1 signals" section suggests that one GWAS was conducted, however the next section compares male and female results, suggesting that a minimum of 3 GWAS were conducted. Why are no results reported separately by sex? Are there any sex differences? If not, maybe there is justification in using the better powered sex-combined results for follow-up analyses. In either case, this should be reported/discussed/justified.
4. How do the results here compare to those reported before? Is it reasonable to double the number of loci identified by increasing the sample size by 20%, or should one be concerned? Of note, there is no QQ plot provided with Figure 1 (normally considered standard reporting).
5. Were there any genetic variants in the IGF1-locus associated with IGF1 levels? A common

criticism is that genetic variants are statistically identified and that biological evidence for their role is needed. If variants in the encoding gene are identified in relation to the circulating protein, this biological positive control adds weight to the validity of the rest of the results.

6. Is it not concerning that IGF1 results correlated so weakly with height, when IGF1 has such an important role in this phenotype? Was this comparison conducted by sex (not specified)? Given that there are strong sex differences in height this seems to be an obvious sanity check.
7. It seems completely inappropriate to be using UKBB to identify genetic variants associated with IGF1, and then to use the same study/individuals to explore the impact of IGF1-associated variants on other phenotypes. All MR analyses in UKBB should be removed.
8. How was multiple testing correction of MR analyses carried out, and what is considered significant in these analyses?
9. What is the rationale for investigating IGF1 in genomic instability? The introduction summarises evidence for IGF1's role in cancer being via apoptosis and proliferation pathways, not via DNA repair or other mechanisms that regulate DNA stability. The hypothesis underlying this test should be clearly stated.
10. If there is rationale for an association between IGF1 and loss of chromosome y in men, should the MR analyses be based on the male-specific GWAS results?
11. MR analyses do not report any assessment of pleiotropy of the genetic variants being used. As lack of pleiotropy is an underlying assumption of MR analyses, this should be presented and discussed. Heidi test and forest plots would also be of value for considering the robustness of findings. Similarly, has overidentification been assessed? Particularly as components of the same pathways are implicated in the MAGENTA/enrichment analyses.
12. Definition and exclusion of "extreme" values of IGF1 seems arbitrary. It is good practice to only exclude data where there are technical issues which explain the values. If this is the case, this should be described and justified.
13. No exclusion criteria are described: should individuals with acromegaly be included? Should individuals with ongoing cancer/cancer therapy at the time of blood sampling be included in the analyses? Should one adjust for conditions such as type 2 diabetes? No effort has been made to explore the phenotypic associations/dependent variables of IGF1, therefore the model used for the association analyses is likely incomplete. Indeed, given the wealth of data available in UKBB, providing these basic descriptive data for IGF1, and conducting genetic analyses of, all ancestry groups would provide a much greater advance in the field than the incremental advance in understanding of "white European" biology provided here.
14. The methods do not mention of whether IGF1 is normally distributed or whether there was any transformation of the data prior to analysis. Again, this is considered standard reporting.

15. The methods for UKBB genetic association analyses used “age and sex” I a model that accounts for population structure and relatedness. The replication analysis included “age, sex, BMI and height” (and study-specific covariates). Why were different models used and can they really be considered “discovery” and “replication” with different models being applied?

Minor comments:

As above, care must be taken when reporting results before methods. Enough methods should be noted so that the results can be interpreted. In addition, the use of abbreviations and where they are defined should be checked.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Expertise in genetic/genomic analyses of complex traits, including GWAS and MR. Familiar with the field of IGF1 biology and with the UK Biobank dataset.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 17 November 2021

<https://doi.org/10.21956/wellcomeopenres.18064.r46972>

© 2021 Janssen J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Joseph AMJL Janssen

Department of Internal Medicine, Erasmus Medical Center, Molewaterplein, Rotterdam, 3015 GD, The Netherlands

Major Comments

1. Previous studies about the role of IGF-I have been inconclusive. The hypothesis of circulating IGF-I being involved in the development of certain cancers is based on the assumption that circulating IGF-I levels are directly related to IGF-I actions at the tissue level. However, it is at present unknown how circulating IGF-I concentrations are related to biological actions of IGF-I at the tissue level. Independent from circulating IGF-I levels, paracrine and autocrine actions of IGF-I may play an important role in the pathogenesis of cancers and IGF-I gene variants may also influence autocrine and paracrine actions of IGF-I in tissues. However, the authors did not consider and discuss this latter possibility.
2. Accurate measurement of IGF-I in serum is crucial. However, there is still a high variability in IGF-I assay performances. Information about the method(s?) (used in the present study and the EPIC-Norfolk study) to measure circulating IGF-I levels is (are) missing.
3. It is assumed that IGF-I only in the free state (i.e. unbound to one of the IGF-BPs) is bioactive and able to bind to the IGF-I receptor. However, in this study only circulating total IGF-I levels were measured.
4. Circulating IGF-I was measured only once. It has been found that the predictive outcomes of earlier studies that used single IGF-I samples for analysis of risk ratios according to tertiles, quartiles, or quintiles could have been different if a second IGF-I was used to establish the risk ratio. These results also suggest that variability in IGF-I should be taken into account when designing such studies. (Milani *et al.* Variability and Reliability of Single Serum IGF-I Measurements: Impact on Determining Predictability of Risk Ratios in Disease Development. *J Clin Endocrinol Metab.* 2004 May; 89(5): 2271–2274). What was the variability of the IGF-I assay (s) used in this study?
5. No information is presented about the relationships between IGF-I levels and cancer in time. When was IGF-I measured? At baseline when subjects were apparently free of cancer? What was the interval between the measurement of IGF-I levels and the actual diagnosis of cancer? When IGF-I levels were measured after cancer diagnosis this may have introduced bias by effects of the disease itself on IGF-I.
6. What was the relationship between IGF variants (IGF-I levels) and stages of cancers at the moment of diagnosis? Were there in this respect differences between early and advanced stages?

References

1. Milani D, Carmichael JD, Welkowitz J, Ferris S, et al.: Variability and reliability of single serum IGF-I measurements: impact on determining predictability of risk ratios in disease development. *J Clin Endocrinol Metab.* 2004; **89** (5): 2271–4 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My main topic of research is the Growth Hormone-Insulin-IGF-I axis in health, aging and disease.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.