





ELSEVIER

## ARTICLE

# Detection and characterization of male sex chromosome abnormalities in the UK Biobank study

Yajie Zhao<sup>1</sup> , Eugene J. Gardner<sup>1</sup>, Marcus A. Tuke<sup>2</sup>, Huairan Zhang<sup>2</sup>, Maik Pietzner<sup>1,3</sup>, Mine Koprulu<sup>1</sup>, Raina Y. Jia<sup>1</sup>, Katherine S. Ruth<sup>2</sup>, Andrew R. Wood<sup>2</sup>, Robin N. Beaumont<sup>2</sup>, Jessica Tyrrell<sup>2</sup>, Samuel E. Jones<sup>2,4</sup>, Hana Lango Allen<sup>1</sup>, Felix R. Day<sup>1</sup>, Claudia Langenberg<sup>1,3</sup>, Timothy M. Frayling<sup>2</sup>, Michael N. Weedon<sup>2</sup>, John R.B. Perry<sup>1</sup>, Ken K. Ong<sup>1,\*</sup> , Anna Murray<sup>2,\*</sup>

## ARTICLE INFO

*Article history:*

Received 31 January 2022

Received in revised form

15 May 2022

Accepted 16 May 2022

Available online xxxx

*Keywords:*

47,XYY

Disorders of sexual development

Endocrinology

Klinefelter syndrome

Thrombosis

Type 2 diabetes

## ABSTRACT

**Purpose:** The study aimed to systematically ascertain male sex chromosome abnormalities, 47,XXY (Klinefelter syndrome [KS]) and 47,XYY, and characterize their risks of adverse health outcomes.

**Methods:** We analyzed genotyping array or exome sequence data in 207,067 men of European ancestry aged 40 to 70 years from the UK Biobank and related these to extensive routine health record data.

**Results:** Only 49 of 213 (23%) of men whom we identified with KS and only 1 of 143 (0.7%) with 47,XYY had a diagnosis of abnormal karyotype on their medical records or self-report. We observed expected associations for KS with reproductive dysfunction (late puberty: risk ratio [RR] = 2.7; childlessness: RR = 4.2; testosterone concentration: RR = -3.8 nmol/L, all  $P < 2 \times 10^{-8}$ ), whereas XYY men appeared to have normal reproductive function. Despite this difference, we identified several higher disease risks shared across both KS and 47,XYY, including type 2 diabetes (RR = 3.0 and 2.6, respectively), venous thrombosis (RR = 6.4 and 7.4, respectively), pulmonary embolism (RR = 3.3 and 3.7, respectively), and chronic obstructive pulmonary disease (RR = 4.4 and 4.6, respectively) (all  $P < 7 \times 10^{-6}$ ).

**Conclusion:** KS and 47,XYY were mostly unrecognized but conferred substantially higher risks for metabolic, vascular, and respiratory diseases, which were only partially explained by higher levels of body mass index, deprivation, and smoking.

Crown Copyright 2022 Published by Elsevier Inc. on behalf of American College of Medical Genetics and Genomics. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Ken K. Ong and Anna Murray jointly supervised this work.

\*Correspondence and requests for materials should be addressed to Ken K. Ong, MRC Epidemiology Unit, Cambridge University School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus Box 285, Cambridge, CB2 0QQ, United Kingdom. *E-mail address:* [Ken.Ong@mrc-epid.cam.ac.uk](mailto:Ken.Ong@mrc-epid.cam.ac.uk) OR Anna Murray, Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, RILD Level 3, Royal Devon & Exeter Hospital, Barrack Road, Exeter, EX2 5DW, United Kingdom. *E-mail address:* [A.Murray@exeter.ac.uk](mailto:A.Murray@exeter.ac.uk)

Affiliations are at the end of the document.

doi: <https://doi.org/10.1016/j.gim.2022.05.011>

1098-3600/Crown Copyright 2022 Published by Elsevier Inc. on behalf of American College of Medical Genetics and Genomics. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

The most common sex chromosome aneuploidies in men are 47,XXY (Klinefelter syndrome [KS]) and 47,XYY, with population prevalence estimates of 100 per 100,000 men and 18 to 100 per 100,000 men,<sup>1,2</sup> respectively. Men with KS typically present during adolescence with delayed puberty or as adults with infertility. Other recognized features include tall adult stature; high body fat percentage;<sup>3</sup> poor muscle tone; low bone mineral density (BMD); and increased risks of neurocognitive disability, psychoses, and disorders of personality.<sup>4</sup> KS has also been associated with higher risks of type 2 diabetes and venous thromboembolism.<sup>5,6</sup> By contrast, 47,XYY is less well-characterized because many of these individuals may not present to health services and thus are unaware of their karyotype. Reported features associated with 47,XYY may therefore be affected by sampling bias. These include tall stature; scoliosis; learning difficulties;<sup>7</sup> poor muscle tone;<sup>8</sup> increased central fat; and increased risks of seizures, asthma, and emotional and behavioral problems (eg, autism and attention deficit disorder).<sup>9</sup> Although infertility has been reported in some men with XYY, most studies report normal sexual development and fertility.<sup>10</sup>

Previous studies identified men with KS or 47,XYY from medical records, and therefore case ascertainment was based on recognition of their typical phenotypic features. Therefore, the reported penetrance of these features may have been biased and the full spectrum of clinical features are overlooked. A more robust alternative approach is to identify such individuals from large population-based studies using systematic measurements to produce unbiased estimates of the effects of sex chromosome aneuploidy on unselected diseases. We recently used this approach to show that mosaic X-chromosome aneuploidy in women (mosaic Turner syndrome, 45,X) conferred a lower penetrance of infertility than that reported by earlier clinic-based studies.<sup>11</sup>

In this study, we analyzed single-nucleotide variation (SNV) array genotype data in 207,067 men of European ancestry aged 40 to 70 years from the UK Biobank. We identified 213 men with sex chromosome aneuploidy indicative of KS and 143 men with 47,XYY and related these karyotypes to extensive study data and medical records to understand the penetrance of male sex chromosome aneuploidy on typical reproductive outcomes and its wider clinical impacts.

## Materials and Methods

### Study population

The UK Biobank is a large prospective cohort that recruited approximately 500,000 participants aged 40 to 70 years across the island of Great Britain. A broad range of phenotypic and health-related information was collected from each participant, including physical measurements,

lifestyle indicators, biomarkers in blood and urine, imaging, and routine health record data.<sup>12</sup>

In the UK Biobank, 488,377 participants had DNA samples assayed using 1 of 2 genotyping arrays: UK Biobank Lung Exome Variant Evaluation (UK BiLEVE study,  $N = 49,950$ ) and Affymetrix Axiom UK Biobank array (UK Biobank Axiom [Affymetrix],  $N = 438,427$ ). These 2 arrays tested 807,411 and 825,927 SNVs, respectively, with 95% overlap between arrays.<sup>12</sup> We restricted our analysis to men of White European genetic ancestry as classified by the approach previously described by Thompson et al.<sup>13</sup> In brief, this approach uses k-means-clustering to group individuals by the first 4 genetic principal components. In addition, we excluded individuals who were classified as White European by our k-means approach but self-identified as being of ancestry other than White European.<sup>13</sup> We further excluded individuals whose samples failed UK Biobank genotyping quality control (QC) parameters and those who withdrew consent. Accordingly, 207,067 men were included in all association testing analyses. We were unable to incorporate non-European individuals when modeling the relationship between abnormal karyotypes and phenotypes outlined in this manuscript. We identified only 16 non-White European males with abnormal karyotypes: 9 with 47,XXY and 7 with 47,XYY.

### Identification of male sex chromosome aneuploidy heterozygotes from SNV array data

To identify men with sex chromosome aneuploidy, we downloaded genotyping fluorescence signal intensity ( $\log_2$  ratios, LRR) and QC information for all SNVs on the X chromosome (chrX) and Y chromosomes (chrY) from the UK Biobank data showcase (<https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=22431> and <https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=1955>). We excluded SNVs that (1) were located within pseudoautosomal regions (PAR),<sup>14</sup> (2) did not have a calculable LRR on both arrays, (3) did not pass QC in all 106 batches, or (4) were flagged as failing QC by UK Biobank. After these steps, 16,599 chrX SNVs and 579 chrY SNVs remained. We then calculated the median LRR across all remaining SNVs on chrX and chrY to generate the values mLRR-X and mLRR-Y, respectively. These values represent the median fluorescence signal intensities across the entire X or Y chromosome.<sup>15</sup> Using the thresholds described by Bycroft et al,<sup>12</sup> men with  $-1 \leq \text{mLRR-Y} < 0.23$  and  $\text{mLRR-X} > -0.2$  were categorized as having 47,XXY (KS) and men with  $\text{mLRR-Y} \geq 0.23$  and  $\text{mLRR-X} < -0.2$  as 47,XYY (men with  $\text{mLRR-Y} \geq 0.23$  and  $\text{mLRR-X} > -0.2$  were categorized as 48,XXYY and were not included in further analyses).

### Confirmation of male sex chromosome aneuploidy heterozygotes from exome sequencing data

To confirm sex chromosome aneuploidy status using an orthogonal approach, we used exome sequencing data

available for 83,104 White European men in UK Biobank.<sup>16,17</sup> To estimate sex chromosome dosage, we calculated the average read depth of 3 target regions: (1) non-PAR regions on chrX, (2) X-degenerate regions (XDRs) on chrY as defined by Skov et al,<sup>18</sup> and (3) autosomes. First, we used SAMtools<sup>19</sup> (version: 1.9) to convert the provided CRAM<sup>20</sup> files for each participant to Binary Alignment Map files on the basis of the GRCh38 reference sequence ([ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/GRCh38\\_reference\\_genome/GRCh38\\_full\\_analysis\\_set\\_plus\\_decoy\\_hla.fa](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/GRCh38_reference_genome/GRCh38_full_analysis_set_plus_decoy_hla.fa)). UK Biobank provided the GRCh38 coordinates of the targeted regions for its exome sequencing design with a Browser Extensible Data file ([https://biobank.ndph.ox.ac.uk/showcase/ukb/auxdata/xgen\\_plus\\_spikein.GRCh38.bed](https://biobank.ndph.ox.ac.uk/showcase/ukb/auxdata/xgen_plus_spikein.GRCh38.bed)). We created 3 subsets of this Browser Extensible Data file by extracting the overlap between the target regions and non-PAR regions on chrX, XDRs on chrY, and autosomes according to their GRCh38 coordinates. Then, they were converted to picard interval lists using the Picard (version: 2.21.6-SNAPSHOT) function `BedToIntervalList`, on the basis of the same reference sequence. Using these picard interval lists, the Binary Alignment Map file of each participant was inputted to calculate the average coverages of non-PAR regions on chrX, XDRs on chrY, and autosomes using the Picard<sup>21</sup> function `CollectHsMetrics`. The relative read depth of non-PAR regions on chrX and XDRs on chrY were defined as the average coverage in each of these regions divided by the average coverage across the autosomes. The relative read depth of non-PAR regions on chrX and XDRs on chrY multiplied by 2 were used as a proxy of chrX dosage and chrY dosage, respectively. Men with chrX dosage > 1.2 were categorized as having 47,XXY (KS) and men with chrY dosage > 1.5 as 47,XYY. Men with chrX dosage > 1.2 and chrY dosage > 1.5 were categorized as 48,XXYY.

## Disease association testing

To test for the disease burden associated with male sex chromosome aneuploidies, we performed logistic regression models with KS or 47,XYY (coded 1) compared with the normal male karyotype 46,XY (coded 0) as the exposure. Outcomes comprised 875 International Classification of Diseases (ICD)-10 coded diseases amalgamated from death registries, hospital episode statistics, primary care records (in a subset,  $n = 94,959$ ), and self-reported conditions (from the first occurrence of disease data set released by UK Biobank). The data set contains further 19 case definitions from dedicated working groups that used multiple sources for case identifications, such as for chronic obstructive pulmonary disease (COPD) or end stage renal disease. For each participant, all events from all sources were mapped to an ICD-10 code and the date of the first disease occurrence from any source was taken as the event date. From this data set, we filtered out likely erroneous disease events if the

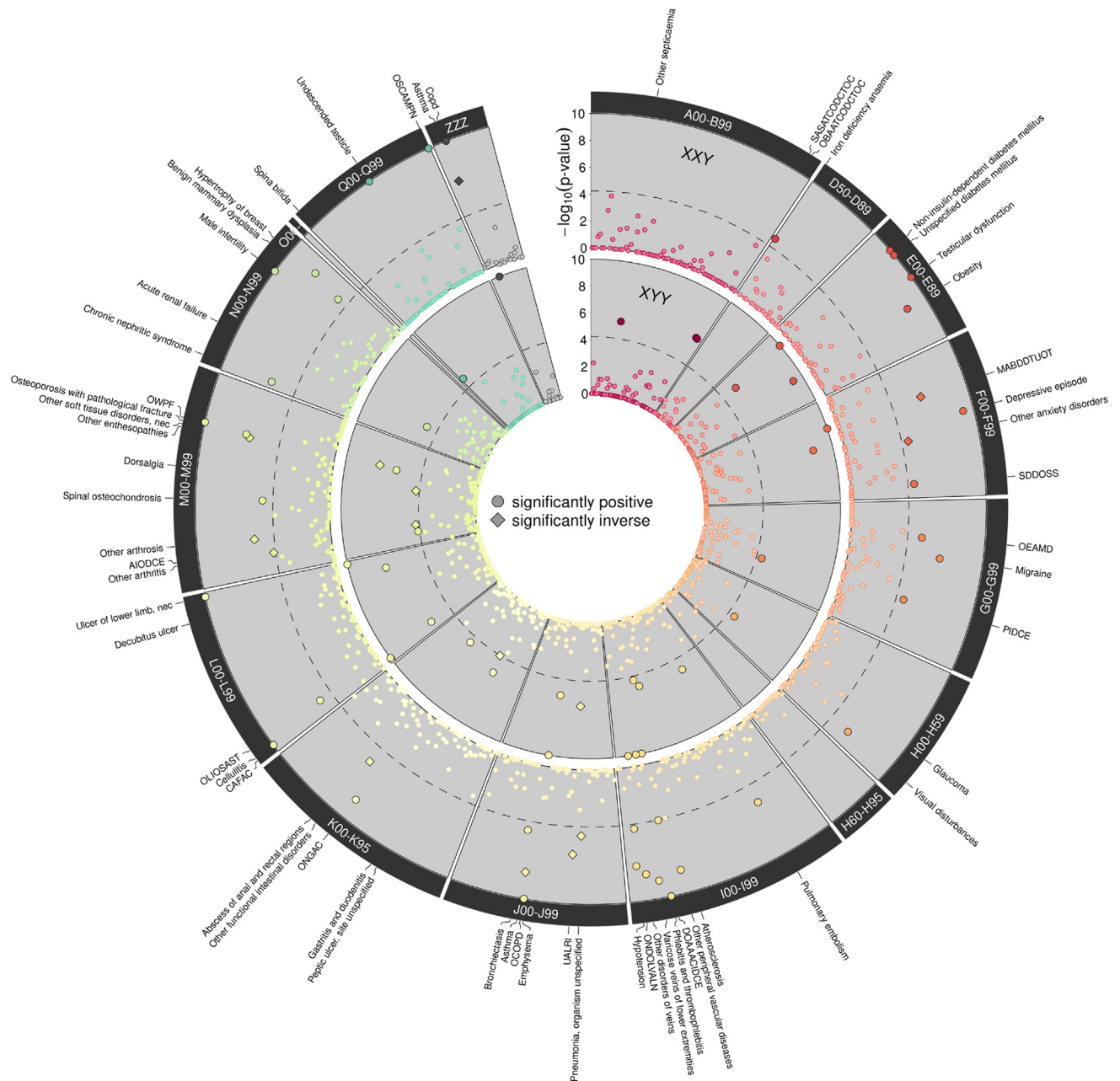
disease occurrence date: (1) was unknown or missing, (2) matched or preceded the date or year of birth, or (3) occurred after the data set release date. We performed logistic regression models in R (version: 3.6.0) among unrelated men of White European genetic ancestry (maximum  $N = 162,322$ ) and adjusted for age at study baseline, test center, and the first 10 genetically derived principal components. Resulting odds ratios were converted to risk ratios (RR) using the formula described by Zhang and Yu.<sup>22</sup> We applied a stringent Bonferroni corrected  $P$  value threshold of  $P < .05/875 = 5.7 \times 10^{-5}$  to define statistical significance (Supplemental Table 1 and Figure 1).

## Study phenotype association testing

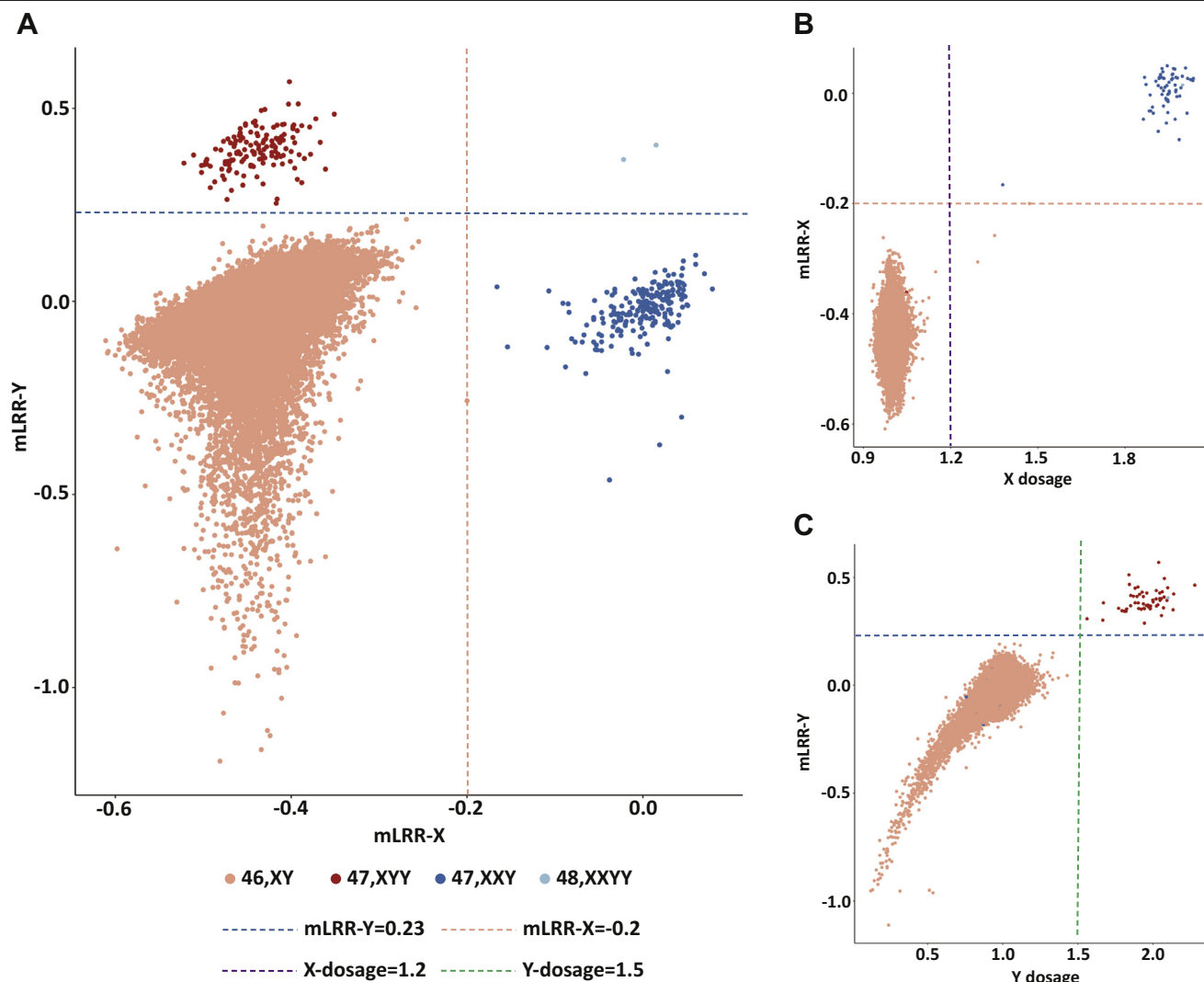
To test the association of male sex chromosome aneuploidy status against selected anthropometric, reproductive, metabolic, cardiovascular, learning/memory, and behavioral study-measured traits (Supplemental Table 2), we used a linear mixed model implemented in BOLT-LMM (version: 2.3.2).<sup>23</sup> The outcome childlessness was derived from the response 0 to the question “How many children have you fathered?” among men aged 55 and older. The Townsend deprivation index was used as an indicator of socioeconomic status, on the basis of participants’ home postcodes. The 2 binary exposure variables described earlier were converted to BGEN file format using `plink2`<sup>24</sup> (version: 2.00-alpha) and inputted to BOLT-LMM via the `bgenFile` flag. A genetic relationship matrix was generated on the basis of all autosomal variants that had minor allele frequency of >1%, passed QC in all 106 batches, and were present on both genotyping arrays. Genotyping chip, age at baseline, and the first 10 genetically derived principal components were included as covariates. For binary outcomes, we also performed logistic regression and calculated the RR from the odds ratio as described earlier.

## Nuclear magnetic resonance metabolic biomarkers association testing

We analyzed 168 circulating metabolic traits measured by proton nuclear magnetic resonance spectroscopy (Nightingale Health Plc) in nonfasting plasma samples in UK Biobank men with 46,XY ( $n = 49,806$ ), KS ( $N = 48$ ), or 47,XYY ( $n = 38$ ). For each metabolic traits, we first performed adjustment for technical variations using the R package `ukbnmr` (unpublished data: Ritchie SC et al 2021. <https://www.medrxiv.org/content/10.1101/2021.09.24.21264079v2>), then performed inverse rank normalization, and then further adjusted for sex, age at the first study visit, body mass index (BMI), and the first 10 genetically derived principal components. Associations between abnormal karyotype and each metabolic trait were tested in separate linear regression models (Supplemental Figure 1).



**Figure 1** Circos plot summarizing genome-wide disease association tests for KS and 47,XXY compared with 46,XY. Each segment represents each International Classification of Diseases (ICD)-10 chapter in lexicographical order.  $P$  values (on a negative logarithmic scale) were from logistic regression models for KS (outer circle) and 47,XXY (inner circle) with each of 875 ICD-10 coded disease outcomes, adjusted for age and 10 principal genetic components. Outcomes reaching the multiple testing corrected statistical significance threshold ( $P < .05/875 = 5.7 \times 10^{-5}$ ; dashed line) are indicated by large circles (for positive associations) and diamonds (for negative associations). AIODCE, arthropathies in other diseases classified elsewhere; CAFAC, cutaneous abscess, furuncle and carbuncle; DOAAACIDCE, disorders of arteries, arterioles and capillaries in diseases classified elsewhere; MABDDTUOT, mental and behavioral disorders due to use of tobacco; NEC, not elsewhere classified; OBAATCODCTOC, other bacterial agents as the cause of diseases classified to other chapters; OCOPD, other chronic obstructive pulmonary disease; OEAMD, other extrapyramidal and movement disorders; ONDOLVALN, other noninfective disorders of lymphatic vessels and lymph nodes; ONGAC, other noninfective gastroenteritis and colitis; OLIOSAST, other local infections of skin and subcutaneous tissue; OSCAMPN, other sex chromosome abnormalities, male phenotype, NEC; OWP, osteoporosis without pathological fracture; PIDCE, polyneuropathy in diseases classified elsewhere; SASATCODCTOC, streptococcus and staphylococcus as the cause of diseases classified to other chapters; SDDOSS, specific developmental disorders of scholastic skills; UALRI, unspecified acute lower respiratory infection.



**Figure 2** Estimates of sex chromosome intensity and dosage by array genotyping or exome sequencing in UK Biobank men. A. Median array genotype intensity on the X chromosome (mLRR-X) and Y chromosome (mLRR-Y) for each of  $n = 207,067$  men, including 213 with 47,XXY (Klinefelter syndrome), 143 with 47,XYY, and 2 with 48,XXYY. B. X dosage estimated from exome sequencing plotted against mLRR-X ( $n = 83,104$ ). C. Y dosage estimated from exome sequencing plotted against mLRR-Y ( $n = 83,104$ ).

## Results

### Prevalence of male sex chromosome aneuploidy in a population scale biobank

Using genotyping array data on 207,067 men of European ancestry, we identified 213 men with 47,XXY (KS, prevalence 103/100,000 men) and 143 with 47,XYY (69/100,000 men; [Figure 2A](#)). Of these cases, who also had exome sequencing data, we observed 100% confirmation of aneuploidy status (62/62 men with KS and 54/54 men with 47,XYY) ([Figure 2B and C](#)).

Only 49 of 213 (23.0%) men with KS and 1 of 143 (0.7%) with 47,XYY had a diagnosis of sex chromosome abnormality on routine medical records or self-reported data (ICD10: Q98 other sex chromosome abnormalities, male phenotype, not elsewhere classified). Similar proportions were found in the subsample of men who had primary care

data: only 24 of 89 (27.0%) with KS and 1 of 76 (1.3%) with 47,XYY had known sex chromosome abnormality. Conversely, of the men with a diagnosis of sex chromosome abnormality on their health record, from our analysis we classified 4 as 46,XX (mLRR-Y < -1) and a further 8 as having a normal male karyotype.

### Quantification of typical features of 47,XXY and 47,XYY

Compared with men with a normal karyotype (46,XY), men with KS and 47,XYY had taller adult height, by 2.7 cm ( $P = 7 \times 10^{-15}$ ) and 7.9 cm ( $P = 8 \times 10^{-77}$ ), respectively, and were more likely to have had taller than average childhood height (RR = 1.3,  $P = .01$  and RR = 1.7,  $P = 5 \times 10^{-10}$ , respectively) ([Table 1](#)). Men with KS and 47,XYY were more likely to be childless (RR = 4.2,  $P = 4 \times 10^{-117}$  and RR = 2.4,  $P = 2 \times 10^{-17}$ , respectively), but only those

**Table 1** Typical features of Klinefelter syndrome and 47,XYY compared with men with normal (46,XY) karyotypes

Features	46,XY			KS			47,XYY			KS vs 46,XY			47,XYY vs 46,XY		
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Beta	95% CI	P Value	Beta	95% CI	P Value	Beta	95% CI	P Value
<b>Continuous traits</b>															
Height, cm	175.9 (6.8)	178.7 (7.6)	184.4 (7.6)	2.7	2.0-2.7	$7 \times 10^{-15}$	7.9	7.0-7.9	$8 \times 10^{-77}$						
Fluid intelligence test scores	6.3 (2.2)	4.2 (1.9)	4.7 (1.9)	-2.2	-2.7 to -2.2	$3 \times 10^{-15}$	-1.6	-2.1 to -1.6	$5 \times 10^{-8}$						
Townsend deprivation index	-1.4 (3.0)	0.3 (3.5)	0.2 (3.5)	1.6	1.2-1.6	$1 \times 10^{-15}$	1.6	1.1-1.6	$1 \times 10^{-10}$						
<b>Binary traits</b>															
	%	%	%	RR	95% CI	P Value	RR	95% CI	P Value						
Childless	21.2%	87.6%	51.8%	4.1	3.9-4.3	$4 \times 10^{-117}$	2.4	2.0-2.8	$2 \times 10^{-17}$						
Taller than average childhood height	31.7%	39.9%	55.6%	1.3	1.0-1.5	$1 \times 10^{-2}$	1.7	1.5-2.2	$5 \times 10^{-10}$						
Late puberty timing	6.2%	17.4%	7.3%	2.7	1.8-3.8	$2 \times 10^{-8}$	1.1	0.5-2.0	$8 \times 10^{-1}$						
University or college degree	40.4%	16.3%	20.2%	0.39	0.3-0.6	$5 \times 10^{-8}$	0.50	0.3-0.7	$9 \times 10^{-5}$						
Depressive episodes	9.1%	22.1%	24.5%	2.4	1.9-3.0	$6 \times 10^{-10}$	2.7	2.0-3.5	$3 \times 10^{-9}$						
Ever smoked	51.6%	54.5%	61.0%	1.1	0.95-1.2	$3 \times 10^{-1}$	1.2	1.1-1.4	$8 \times 10^{-3}$						
Lives alone	17.2%	37.4%	35.3%	2.1	1.8-2.5	$3 \times 10^{-14}$	2.1	1.6-2.5	$2 \times 10^{-8}$						
Lives without a partner	22.7%	50.0%	48.6%	2.2	1.9-2.5	$4 \times 10^{-20}$	2.1	1.8-2.5	$1 \times 10^{-12}$						
Loneliness, isolation	14.8%	33.3%	37.5%	2.2	1.8-2.6	$4 \times 10^{-13}$	2.5	2.0-3.0	$3 \times 10^{-13}$						
Poor overall health	5.0%	21.1%	19.1%	4.2	3.2-5.3	$2 \times 10^{-26}$	3.8	2.6-5.2	$2 \times 10^{-14}$						
Long-standing illness or infirmity	35.7%	65.7%	63.8%	1.9	1.7-2.1	$1 \times 10^{-19}$	1.8	1.6-2.1	$4 \times 10^{-13}$						

Beta, regression coefficient from linear regression models; KS, Klinefelter syndrome; RR, Relative risk from logistic regression models.

with KS (not 47,XYY) were more likely to report later than average puberty timing (RR = 2.7,  $P = 2 \times 10^{-8}$ ).

In addition, men with KS and 47,XYY were less likely to have a university or college degree (RR = 0.39,  $P = 5 \times 10^{-8}$  and RR = 0.50,  $P = 9 \times 10^{-5}$ , respectively) and had lower fluid intelligence test scores (beta = -2.1,  $P = 3 \times 10^{-15}$  and beta = -1.6,  $P = 5 \times 10^{-8}$ , respectively), were more likely to be smokers (RR = 1.1,  $P = 0.28$  and RR = 1.2,  $P = 8 \times 10^{-3}$ , respectively), report depressive episodes (RR = 2.4,  $P = 6 \times 10^{-10}$  and RR = 2.7,  $P = 3 \times 10^{-9}$ , respectively), and live in areas with higher deprivation index (beta = 1.6,  $P = 1 \times 10^{-15}$  and beta = 1.6,  $P = 1 \times 10^{-10}$ , respectively). Men with KS and 47,XYY were also more likely to live without a partner (RR = 2.2,  $P = 4 \times 10^{-20}$  and RR = 2.1,  $P = 1 \times 10^{-12}$ , respectively) and report loneliness and isolation (RR = 2.2,  $P = 4 \times 10^{-13}$  and RR = 2.5,  $P = 3 \times 10^{-13}$ , respectively) and poor overall health (RR = 4.2,  $P = 2 \times 10^{-26}$  and RR = 3.8,  $P = 2 \times 10^{-14}$ , respectively).

### Anthropometric features of men with 47,XXY and 47,XYY

Compared with 46,XY men, those with KS and 47,XYY had higher BMI (beta = 1.8 kg/m<sup>2</sup>,  $P = 8 \times 10^{-11}$  and beta = 2.2 kg/m<sup>2</sup>,  $P = 5 \times 10^{-11}$ , respectively), higher percentage of total body fat (beta = 4.8%,  $P = 8 \times 10^{-40}$  and beta = 2.2%,  $P = 3 \times 10^{-7}$ , respectively), and weaker hand grip strength (beta = -7.1 kg,  $P = 1 \times 10^{-36}$  and beta = -2.6 kg,  $P = 3 \times 10^{-4}$ , respectively) (Table 2). Other features differed between the groups. Even after accounting for their taller adult heights, men with 47,XYY (but not KS) had higher total fat-free mass (beta = 3.8 Kg,  $P = 4 \times 10^{-15}$ ). Furthermore, men with 47,XYY had slightly higher BMD (beta = 0.04 g/cm<sup>2</sup>,  $P = 2 \times 10^{-2}$ ), whereas men with KS had lower BMD (beta = -0.05 g/cm<sup>2</sup>,  $P = 5 \times 10^{-6}$ ) and

higher likelihood of osteoporosis with pathological fractures (RR = 10.8,  $P = 2 \times 10^{-7}$ ).

### Hormonal, metabolic, and vascular features of men with 47,XXY and 47,XYY

Compared with 46,XY men, those with KS (but not 47,XYY) had lower plasma total testosterone concentration (beta = -3.8 nmol/L,  $P = 2 \times 10^{-50}$ ). Both KS and 47,XYY men had lower plasma IGF-1 concentrations (beta = -1.7 nmol/L,  $P = 1 \times 10^{-6}$  and beta = -2.3,  $P = 3 \times 10^{-8}$ , respectively).

We also identified several adverse metabolic and vascular health outcomes shared across both KS and 47,XYY, including higher risks for type 2 diabetes (RR = 3.0,  $P = 2 \times 10^{-20}$  and RR = 2.6,  $P = 3 \times 10^{-10}$ , respectively), albuminuria (RR = 1.9,  $P = 5 \times 10^{-4}$  and RR = 2.4,  $P = 5 \times 10^{-6}$ , respectively), venous thrombosis (RR = 6.4,  $P = 3 \times 10^{-23}$  and RR = 7.4,  $P = 7 \times 10^{-22}$ , respectively), pulmonary embolism (RR = 3.3,  $P = 2 \times 10^{-6}$  and RR = 3.7,  $P = 7 \times 10^{-6}$ , respectively), and atherosclerosis (RR = 3.1,  $P = 6 \times 10^{-3}$  and RR = 5.5,  $P = 8 \times 10^{-6}$ , respectively). These disease associations were only partially attenuated after adjustments for BMI, household deprivation, and smoking (Table 3). Exploration of red blood cell and platelet traits showed lower hematocrit and hemoglobin concentrations in KS (but not 47,XYY) men, but no obvious explanation for higher thrombosis risk was found (Supplemental Table 2).

Men with KS and 47,XYY had lower levels of high-density lipoprotein (HDL) cholesterol (beta = -0.11 mmol/L,  $P = 2 \times 10^{-7}$  and beta = -0.17 mmol/L,  $P = 3 \times 10^{-11}$ , respectively) and men with 47,XYY (but not KS) had higher triglycerides (beta = 0.32 mmol/L,  $P = 1 \times 10^{-3}$ ). In the subgroup with nuclear magnetic resonance metabolic data, we observed that KS ( $n = 48$ ) and 47,XYY ( $n = 38$ ) men had lower levels of most HDL-related traits. Furthermore,

**Table 2** Anthropometric characteristics of Klinefelter syndrome and 47,XYY compared with men with normal (46,XY) karyotypes

Characteristics	46,XY	KS	47,XYY	KS vs 46,XY			47,XYY vs 46,XY		
	Mean (SD)	Mean (SD)	Mean (SD)	Beta	95% CI	P Value	Beta	95% CI	P Value
<b>Continuous traits</b>									
BMI, kg/m <sup>2</sup>	27.9 (4.2)	29.7 (5.7)	30.2 (5.7)	1.8	1.2-2.3	8 × 10 <sup>-11</sup>	2.2	1.6-2.9	5 × 10 <sup>-11</sup>
Body fat percentage, %	29.5 (5.4)	34.3 (6.0)	31.6 (5.7)	4.8	4.1-5.6	9 × 10 <sup>-40</sup>	2.2	1.4-3.1	4 × 10 <sup>-7</sup>
Hand grip strength, kg	41.9 (8.9)	35.2 (8.9)	39.9 (9.7)	-7.1	-8.2 to -6.0	1 × 10 <sup>-36</sup>	-2.6	-3.9 to -1.2	3 × 10 <sup>-4</sup>
Fat-free mass, adj. height, kg	63.9 (7.7)	66.1 (9.4)	73.5 (9.2)	0.2	-0.6 to 0.9	6 × 10 <sup>-1</sup>	3.8	2.8-4.7	4 × 10 <sup>-15</sup>
Bone mineral density, g/cm <sup>2</sup>	0.6 (0.2)	0.5 (0.2)	0.6 (0.2)	-0.05	-0.08 to -0.03	5 × 10 <sup>-6</sup>	0.04	0.01-0.08	2 × 10 <sup>-2</sup>
<b>Binary traits</b>	%	%	%	RR	95% CI	P Value	RR	95% CI	P Value
Osteoporosis with pathological fracture	0.3%	2.4%	0.7%	10.8	4.5-25.3	2 × 10 <sup>-7</sup>	3.1	0.4-20.6	3 × 10 <sup>-1</sup>
Osteoporosis without pathological fracture	1.6%	8.9%	0.7%	6.2	4.0-8.9	5 × 10 <sup>-15</sup>	0.5	0.1-3.4	5 × 10 <sup>-1</sup>

Beta, regression coefficient from linear regression models; BMI, body mass index; KS, Klinefelter syndrome; RR, relative risk from logistic regression models.

47,XYY (but not KS) men showed lower levels across all lipid classes, apart from triglycerides (Supplemental Figure 1).

### Respiratory features of men with 47,XYY and 47,XYY

Compared with 46,XY men, those with KS and 47,XYY had lower forced expiratory volume (beta = -0.68 L,  $P = 1 \times 10^{-46}$  and beta = -0.26 L,  $P = 2 \times 10^{-5}$ , respectively), peak expiratory flow (beta = -125 L/min,  $P = 2 \times 10^{-38}$  and beta = -68 L/min,  $P = 4 \times 10^{-8}$ , respectively) and vital capacity (beta = -0.93 L,  $P = 2 \times 10^{-50}$  and beta = -0.25 L,  $P = 3 \times 10^{-3}$ , respectively), and higher risks for COPD (RR = 4.4,  $P = 5 \times 10^{-18}$  and RR = 4.6,  $P = 2 \times 10^{-13}$ , respectively) and asthma (RR = 2.0,  $P = 9 \times 10^{-9}$  and RR = 1.7,  $P = 4 \times 10^{-3}$ , respectively). Again, these disease associations were only partially attenuated after adjustments for BMI, household deprivation, and smoking (Table 3).

### Discussion

Using systematic case ascertainment in a large, unselected population of men of European ancestry aged 40 to 70 years, we report the prevalence of KS (103/100,000 men) and 47,XYY (69/100,000 men). Notably, only a small minority of these men had a diagnosis of sex chromosome abnormality on their medical records or by self-report (23% of KS and 0.7% of 47,XYY) and yet these conditions conferred substantially increased risks for multiple, potentially preventable diseases.

The underdiagnosis of KS and 47,XYY has been previously indirectly quantified in other settings, on the basis of the differences in their clinical prevalence compared with estimates from population-based cytogenetic surveys in newborn infants. Such studies estimated that only between 7% (in the United Kingdom) and 57% (in Australia) of

expected KS cases were diagnosed on the basis of clinical presentation. For XYY, only between 3% (in the United Kingdom) and 18% (in Denmark) of expected cases were diagnosed.<sup>2</sup>

Our prevalence estimates in an adult study population are somewhat lower than those reported in those newborn infants (KS: 152/100,000 males and 47,XYY: 98/100,000 males).<sup>2</sup> Although this could be interpreted as indicating higher mortality rates, it is recognized that UK Biobank comprises a more educated and healthier sample than the general population, likely owing to healthy volunteer bias.<sup>25</sup> Similarly, the prevalence of other adverse genetic conditions is reportedly lower in UK Biobank than in other more representative studies.<sup>26</sup>

Previous studies have highlighted higher disease risks in men with KS. Bojesen et al<sup>27</sup> identified 832 men with KS from hospital records in Denmark and reported higher risks for venous thrombosis (hazard ratio = 5.3, 95% CI = 3.3-8.5), pulmonary embolism (hazard ratio = 3.6, 95% CI = 1.9-6.7), COPD (hazard ratio = 3.9, 95% CI = 2.5-6.1), type 2 diabetes (hazard ratio = 3.7, 95% CI = 2.1-6.4), and atherosclerosis (hazard ratio = 4.5, 95% CI = 2.8-7.1). Swerdlow et al<sup>6</sup> accessed data on 3518 UK patients with KS diagnosed since 1959 and followed up till mid-2003 and reported higher mortality from diabetes mellitus (standardized mortality ratio = 5.8, 95% CI = 3.4-9.3), pulmonary embolism (standardized mortality ratio = 5.7, 95% CI = 2.5-11.3), and chronic lower respiratory disease (standardized mortality ratio = 2.1, 95% CI = 1.4-3.0). Zöller et al<sup>5</sup> reported a higher risk for venous thromboembolism (incidence rate ratio [IRR] = 6.4, 95% CI = 5.1-7.9) in 1085 men diagnosed with KS between 1969 and 2010 in Sweden. Our findings confirm these strong disease associations and also the reported higher risks of psychiatric illness and osteoporosis.

By contrast, few studies have reported on disease risks in men with 47,XYY. Berglund et al<sup>28</sup> identified 255 men with 47,XYY from hospital records in Denmark and reported higher risks for venous thrombosis (IRR = 10.2, 95% CI = 4.6-22.6) and COPD (IRR = 5.8, 95% CI = 2.4-15.1). Our

**Table 3** Hormonal, metabolic, vascular, and respiratory characteristics of men with Klinefelter syndrome and 47,XYY compared with men with normal (46,XY) karyotypes

Characteristics	46,XY	KS	47,XYY	KS vs 46,XY						47,XYY vs 46,XY					
				Baseline model			Adj. BMI, TDI, and smoking			Baseline model			Adj. BMI, TDI, and smoking		
Continuous traits	Mean (SD)	Mean (SD)	Mean (SD)	Beta	95% CI	P Value	Beta	95% CI	P Value	Beta	95% CI	P Value	Beta	95% CI	P Value
Testosterone, nmol/L	12.0 (3.7)	8.2 (6.0)	11.5 (4.7)	-3.8	-4.3 to -3.3	$2 \times 10^{-50}$	-3.3	-3.8 to -2.8	$2 \times 10^{-41}$	-0.5	-1.1 to 0.1	$9 \times 10^{-2}$	0.1	-0.5 to 0.7	$8 \times 10^{-1}$
SHBG, nmol/L	39.9 (16.8)	41.1 (22.7)	40.3 (20.4)	2.3	0.1-4.4	$2 \times 10^{-2}$	4.2	2.1-6.2	$3 \times 10^{-5}$	1.3	-1.4 to 4.0	$3 \times 10^{-1}$	3.7	1.1-6.2	$2 \times 10^{-3}$
IGF-1, nmol/L	21.9 (5.5)	20.4 (6.5)	20.0 (5.7)	-1.7	-2.3 to -1.0	$1 \times 10^{-6}$	-1.2	-1.9 to -0.5	$7 \times 10^{-4}$	-2.3	-3.1 to -1.4	$3 \times 10^{-8}$	-1.8	-2.6 to -0.9	$2 \times 10^{-5}$
HDL cholesterol, mmol/L	1.3 (0.3)	1.2 (0.3)	1.1 (0.3)	-0.11	-0.15 to -0.07	$2 \times 10^{-7}$	-0.06	-0.10 to -0.02	$5 \times 10^{-3}$	-0.17	-0.22 to -0.12	$3 \times 10^{-11}$	-0.10	-0.15 to -0.06	$1 \times 10^{-5}$
LDL cholesterol, mmol/L	3.5 (0.9)	3.4 (0.9)	3.4 (0.9)	-0.10	-0.22 to 0.01	$6 \times 10^{-2}$	-0.06	-0.18 to 0.05	$3 \times 10^{-1}$	-0.15	-0.29 to -0.01	$2 \times 10^{-2}$	-0.11	-0.25 to 0.03	$9 \times 10^{-2}$
Triglycerides, mmol/L	2.0 (1.2)	2.1 (1.0)	2.3 (1.4)	0.10	-0.05 to 0.25	$2 \times 10^{-1}$	-0.01	-0.16 to 0.13	$8 \times 10^{-1}$	0.32	0.14-0.51	$1 \times 10^{-3}$	0.13	-0.05 to 0.31	$3 \times 10^{-1}$
Forced expiratory volume, L	3.3 (0.8)	2.7 (0.8)	3.2 (0.9)	-0.7	-0.8 to -0.6	$1 \times 10^{-46}$	-0.6	-0.7 to -0.5	$1 \times 10^{-39}$	-0.3	-0.4 to -0.1	$2 \times 10^{-5}$	-0.2	-0.3 to -0.1	$5 \times 10^{-3}$
Peak expiratory flow, L/min	469 (138)	348 (132)	408 (132)	-125	-144 to -106	$2 \times 10^{-38}$	-117	-136 to -98	$3 \times 10^{-34}$	-68	-92 to -44	$4 \times 10^{-8}$	-58	-81 to -34	$4 \times 10^{-6}$
Vital capacity, L	4.5 (1.0)	3.5 (0.9)	4.3 (1.0)	-0.9	-1.1 to -0.8	$2 \times 10^{-50}$	-0.8	-0.9 to -0.7	$4 \times 10^{-41}$	-0.2	-0.4 to -0.1	$3 \times 10^{-3}$	-0.1	-0.03 to 0.03	$2 \times 10^{-1}$
Binary traits	%	%	%	RR	95% CI	P Value	RR	95% CI	P Value	RR	95% CI	P Value	RR	95% CI	P Value
Type 2 diabetes	10.3%	28.6%	25.0%	3.0	2.4-3.6	$2 \times 10^{-20}$	2.3	1.7-2.9	$5 \times 10^{-12}$	2.6	1.9-3.4	$3 \times 10^{-10}$	1.8	1.2-2.5	$3 \times 10^{-5}$
Albuminuria	13.9%	26.6%	30.3%	1.9	1.3-2.6	$5 \times 10^{-4}$	1.7	1.2-2.4	$4 \times 10^{-3}$	2.4	1.7-3.2	$5 \times 10^{-6}$	2.1	1.4-2.9	$1 \times 10^{-4}$
Venous thrombosis	1.8%	10.9%	12.6%	6.4	4.2-9.1	$3 \times 10^{-23}$	5.4	3.5-7.8	$2 \times 10^{-22}$	7.4	4.6-10.9	$7 \times 10^{-22}$	5.6	3.4-8.6	$2 \times 10^{-18}$
Pulmonary embolism	2.5%	4.9%	7.5%	3.3	2.0-5.2	$2 \times 10^{-6}$	2.8	1.7-4.5	$5 \times 10^{-5}$	3.7	2.1-6.2	$7 \times 10^{-6}$	2.8	1.5-4.9	$9 \times 10^{-4}$
Atherosclerosis	1.0%	2.8%	4.9%	3.1	1.4-6.8	$6 \times 10^{-3}$	2.2	0.9-5.2	$9 \times 10^{-2}$	5.5	2.6-11.2	$8 \times 10^{-6}$	4.4	2.1-9.1	$1 \times 10^{-4}$
COPD	4.2%	16.4%	16.8%	4.4	3.2-6.0	$5 \times 10^{-18}$	3.7	2.6-5.1	$2 \times 10^{-12}$	4.6	3.1-6.5	$2 \times 10^{-13}$	3.3	2.1-5.0	$2 \times 10^{-7}$
Asthma	11.8%	23.5%	19.6%	2.0	1.6-2.6	$1 \times 10^{-7}$	1.9	1.5-2.5	$1 \times 10^{-6}$	1.7	1.2-2.3	$4 \times 10^{-3}$	1.6	1.1-2.2	$1 \times 10^{-2}$

Adj, adjusted for; Beta, regression coefficient from linear regression models; BMI, body mass index; COPD, chronic obstructive pulmonary disease; HDL, high-density lipoprotein; KS, Klinefelter syndrome; LDL, low-density lipoprotein; RR, relative risk from logistic regression models; TDI, Townsend deprivation index.



findings confirm those observations and extend the list of diseases strongly associated with 47,XYY to also include type 2 diabetes, pulmonary embolism, and atherosclerosis.

We observed some notable differences between KS and 47,XYY. KS is a well-recognized cause of reproductive dysfunction and this was reflected in our data by their older age at puberty, lower testosterone levels, and high risk of being childless. Reproductive dysfunction likely also contributes to their tall stature (due to later pubertal growth completion), lower bone density and muscle strength, and greater adiposity. By contrast, men with 47,XYY appeared to have normal reproductive function, with no alteration in their puberty timing or testosterone levels and a more modestly higher risk of being childless, which could be explained by their similarly higher chance of living without a partner.

Hence, despite these marked differences in reproductive function, it is unclear why both KS and 47,XYY should show striking similarities in conferring substantially higher risks for many diseases in common—type 2 diabetes, atherosclerosis, venous thrombosis, pulmonary embolism, and COPD, which persisted after adjustments for several lifestyle behavioral-related traits (BMI, smoking, deprivation). Higher risks of type 2 diabetes, atherosclerosis, and microalbuminuria, with lower HDL cholesterol and higher adiposity, together indicate higher insulin resistance in both KS and 47,XYY men. Both conditions confer a triple dose of the PAR, containing the growth-promoting *SHOX* gene, which likely partially contributes to their tall stature, and there are case reports of insulin resistance in other conditions characterized by *SHOX* excess.<sup>29,30</sup> However, the underlying mechanisms are yet unknown.

Similarly, it is unclear why risks for venous thrombosis and pulmonary embolism are raised in both KS and 47,XYY to a similar substantial degree, around 6- to 7-fold higher risk for venous thrombosis. This is similar or even higher than that conferred by factor V Leiden, a genetic variant carried by around 5% of White and of European descent individuals.<sup>31</sup> Hence, it might be considered to add sex chromosome aneuploidy to the screening for genetic causes of thrombophilia. Furthermore, because KS and 47,XYY confer higher risks for multiple potentially preventable diseases, future studies should explore the potential benefits of wider testing.

Strengths of our study include the use of systematic case ascertainment and 100% confirmation of genome-wide association study (GWAS) array-based categorization using exome sequencing data in a large subsample. Furthermore, our numbers of individuals with a male sex chromosome aneuploidy are similar to those reported by UK Biobank, which found 355 such individuals (99.2% of our 358 cases).<sup>12</sup> GWAS array genotyping and exome sequencing are increasingly performed in clinical settings, however male sex chromosome aneuploidy status is not routinely derived. In addition, the wide range of traits and diseases available in UK Biobank allowed us to systematically quantify the disease and phenotypic effects of male sex chromosome aneuploidy.

Limitations include the healthy volunteer bias of the UK Biobank sample and the yet incomplete linkage to primary care health data. Hence it is likely that the true disease risks associated with KS and 47,XYY analyses are even higher than the substantial estimates that we observed.

In conclusion, our findings show that male sex chromosome aneuploidy can be reliably detected using GWAS or exome sequencing data. KS and 47,XYY were mostly unrecognized but conferred substantially higher risks of diverse potentially preventable diseases, including metabolic and vascular diseases, which were only partially explained by higher levels of BMI, deprivation, and smoking. Future studies should consider the utility of deriving male sex chromosome aneuploidy status when genetic testing is undertaken for existing clinical indications, eg, for thrombosis risk. Furthermore, our findings add significantly to ongoing debates regarding the potential benefits of wider population genetic screening.

## Data Availability

All data are available via the UK Biobank Access Management System <https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>.

## Acknowledgments

Y.Z., E.J.G., M.P., M.K., R.Y.J., H.L.A., F.R.D., C.L., J.R.B.P. and K.K.O. are supported by the Medical Research Council, United Kingdom (Unit programs: MC\_UU\_00006/1 and MC\_UU\_00006/2). M.K. is supported by a Gates Fellowship. A.R.W. and J.T. hold Academy of Medical Sciences Springboard awards [SBF004\1079 and SBF006\1134], which are supported by the Academy of Medical Sciences, United Kingdom; the Wellcome Trust; the Government Department of Business, Energy and Industrial Strategy; the British Heart Foundation; and Diabetes UK, United Kingdom. R.N.B. and A.M. are supported by the Medical Research Council (MR/T00200X/1). K.S.R. is supported by Cancer Research UK, United Kingdom [C18281/A29019]. The authors acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work. The funders had no role in the design, analysis, and interpretation of data; writing of the report; or the decision to submit the article for publication.

## Author Information

Conceptualization: J.R.B.P., K.K.O., A.M.; Data Curation: Y.Z., M.P., M.K., R.Y.J., F.R.D.; Formal Analysis: Y.Z., E.J.G., M.A.T., H.Z., M.P., M.K., R.Y.J., K.S.R., A.R.W., R.N.B., J.T., S.E.J.; Funding Acquisition: J.R.B.P., K.K.O., A.M.; Methodology: Y.Z., E.J.G., H.L.A.,

J.R.B.P., K.K.O.; Supervision: C.L., T.M.F., M.N.W., J.R.B.P., K.K.O., A.M.; Visualization: Y.Z., M.P.; Writing-original draft: Y.Z., E.J.G., J.R.B.P., K.K.O., A.M.; Writing-review and editing: Y.Z., E.J.G., M.A.T., H.Z., M.P., M.K., R.Y.J., K.S.R., A.R.W., R.N.B., J.T., S.E.J., H.L.A., F.R.D., C.L., T.M.F., M.N.W., J.R.B.P., K.K.O., A.M.

## Ethics Declaration

UK Biobank has approval from the North West Multi-centre Research Ethics Committee (REC reference 13/NW/0157, <https://www.ukbiobank.ac.uk/media/lcvbdoik/21-nw-0157-favourable-opinion-with-conditions-18-06-2021.pdf>) as a Research Tissue Bank (RTB) approval and informed consent (<https://www.ukbiobank.ac.uk/media/t22hbo35/consent-form.pdf>) was provided by each participant. This approval means that researchers do not require separate ethical clearance and can operate under the RTB approval. This RTB approval was granted initially in 2011 and it is a renewal on a 5-yearly cycle; hence UK Biobank successfully applied to renew it in 2016 and 2021. This research was conducted using the UK Biobank resource under applications 9905 (<https://www.ukbiobank.ac.uk/enable-your-research/approved-research/using-genetics-to-elucidate-the-relationship-between-socioeconomic-behavioural-traits-and-disease>), 9072 (<https://biobank.ctsu.ox.ac.uk/crystal/app.cgi?id=9072>), and 49847 (<https://www.ukbiobank.ac.uk/enable-your-research/approved-research/assessing-the-prevalence-penetrance-and-pathogenicity-of-rare-genomic-variants>).

## Conflict of Interest

The authors declare no conflicts of interest.

## Additional Information

The online version of this article (<https://doi.org/10.1016/j.gim.2022.05.011>) contains supplementary material, which is available to authorized users.

## Affiliations

<sup>1</sup>MRC Epidemiology Unit, Institute of Metabolic Science, School of Clinical Medicine, Cambridge University, Cambridge, United Kingdom; <sup>2</sup>Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Royal Devon & Exeter Hospital, Exeter, United Kingdom; <sup>3</sup>Computational Medicine, Berlin Institute of Health (BIH) at Charité, Universitätsmedizin Berlin, Berlin, Germany; <sup>4</sup>Institute for Molecular Medicine Finland (FIMM), Helsinki

Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland

## References

- Ratcliffe S. Long-term outcome in children of sex chromosome abnormalities. *Arch Dis Child*. 1999;80(2):192–195. <http://doi.org/10.1136/adc.80.2.192>.
- Berglund A, Viuff MH, Skakkebaek A, Chang S, Stochholm K, Gravholt CH. Changes in the cohort composition of turner syndrome and severe non-diagnosis of Klinefelter, 47,XXX and 47,YYY syndrome: a nationwide cohort study. *Orphanet J Rare Dis*. 2019;14(1):16. <http://doi.org/10.1186/s13023-018-0976-2>.
- Chang S, Skakkebaek A, Trolle C, et al. Anthropometry in Klinefelter syndrome – multifactorial influences due to CAG length, testosterone treatment and possibly intrauterine hypogonadism. *J Clin Endocrinol Metab*. 2015;100(3):E508–E517. <http://doi.org/10.1210/jc.2014-2834>.
- Skakkebaek A, Moore PJ, Pedersen AD, et al. The role of genes, intelligence, personality, and social engagement in cognitive performance in Klinefelter syndrome. *Brain Behav*. 2017;7(3):e00645. <http://doi.org/10.1002/brb3.645>.
- Zöller B, Ji J, Sundquist J, Sundquist K. High risk of venous thromboembolism in Klinefelter syndrome. *J Am Heart Assoc*. 2016;5(5):e003567. <http://doi.org/10.1161/JAHA.116.003567>.
- Swerdlow AJ, Higgins CD, Schoemaker MJ, Wright AF, Jacobs PA, United Kingdom Clinical Cytogenetics Group. Mortality in patients with Klinefelter syndrome in Britain: a cohort study. *J Clin Endocrinol Metab*. 2005;90(12):6516–6522. <http://doi.org/10.1210/jc.2005-1077>.
- Leggett V, Jacobs P, Nation K, Scerif G, Bishop DVM. Neurocognitive outcomes of individuals with a sex chromosome trisomy: XXX, XYY, or XXY: a systematic review. *Dev Med Child Neurol*. 2010;52(2):119–129. <http://doi.org/10.1111/j.1469-8749.2009.03545.x>.
- Asano A, Motomura N, Yokota S, Yoneda H, Sakai T, Tsutsumi S. Myotonic dystrophy associated with 47 XYY syndrome. *Psychiatry Clin Neurosci*. 2000;54(1):113–116. <http://doi.org/10.1046/j.1440-1819.2000.00645.x>.
- Tartaglia NR, Wilson R, Miller JS, et al. Autism spectrum disorder in males with sex chromosome aneuploidy: XXY/Klinefelter Syndrome, XYY, and XYY. *J Dev Behav Pediatr*. 2017;38(3):197–207. <http://doi.org/10.1097/DBP.0000000000000429>.
- Borjian Boroujeni P, Sabbaghian M, Vosough Dizaji A, et al. Clinical aspects of infertile 47,XYY patients: a retrospective study. *Hum Fertil (Camb)*. 2019;22(2):88–93. <http://doi.org/10.1080/14647273.2017.1353143>.
- Tuke MA, Ruth KS, Wood AR, et al. Mosaic Turner syndrome shows reduced penetrance in an adult population study. *Genet Med*. 2019;21(4):877–886. <http://doi.org/10.1038/s41436-018-0271-6>.
- Bycroft C, Freeman C, Petkova D, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–209. <http://doi.org/10.1038/s41586-018-0579-z>.
- Thompson DJ, Genovese G, Halvardson J, et al. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature*. 2019;575(7784):652–657. <http://doi.org/10.1038/s41586-019-1765-3>.
- Flaquer A, Rappold GA, Wienker TF, Fischer C. The human pseudoautosomal regions: a review for genetic epidemiologists. *Eur J Hum Genet*. 2008;16(7):771–779. <http://doi.org/10.1038/ejhg.2008.63>.
- 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. <http://doi.org/10.1038/ng.3821>.
- Szastakowski JD, Balasubramanian S, Kvikstad E, et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet*. 2021;53(7):942–948. <http://doi.org/10.1038/s41588-021-00885-0>.

17. Yun T, Li H, Chang PC, Lin MF, Carroll A, McLean CY. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. *Bioinformatics*. 2021;36(24):5582–5589. <http://doi.org/10.1093/bioinformatics/btaa1081>.
18. Skov L, , Danish Pan Genome Consortium, Schierup MH. Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. *PLoS Genet*. 2017;13(8):e1006834. <http://doi.org/10.1371/journal.pgen.1006834>.
19. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008. <http://doi.org/10.1093/gigascience/giab008>.
20. Bonfield JK. CRAM 3.1: Advances in the CRAM File Format [published online ahead of print, 2022 Jan 6]. *Bioinformatics*. 2022;38(6):1497–1503. <http://doi.org/10.1093/bioinformatics/btac010>.
21. Bonfield JK, Marshall J, Danecek P, et al. C library for reading/writing high-throughput sequencing data. *Gigascience*. 2021;10(2):giab007. <http://doi.org/10.1093/gigascience/giab007>.
22. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*. 1998;280(19):1690–1691. <http://doi.org/10.1001/jama.280.19.1690>.
23. Loh PR, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47(3):284–290. <http://doi.org/10.1038/ng.3190>.
24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. <http://doi.org/10.1186/s13742-015-0047-8>.
25. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026–1034. <http://doi.org/10.1093/aje/kwx246>.
26. Wade KH, Lam BYH, Melvin A, et al. Loss-of-function mutations in the melanocortin 4 receptor in a UK birth cohort. *Nat Med*. 2021;27(6):1088–1096. <http://doi.org/10.1038/s41591-021-01349-y>.
27. Bojesen A, Juul S, Birkebaek NH, Gravholt CH. Morbidity in Klinefelter syndrome: a Danish register study based on hospital discharge diagnoses. *J Clin Endocrinol Metab*. 2006;91(4):1254–1260. <http://doi.org/10.1210/jc.2005-0697>.
28. Berglund A, Stochholm K, Gravholt CH. Morbidity in 47,XYY syndrome: a nationwide epidemiological study of hospital diagnoses and medication use. *Genet Med*. 2020;22(9):1542–1551. <http://doi.org/10.1038/s41436-020-0837-y>.
29. Kanaka-Gantenbein C, Kitsiou S, Mavrou A, et al. Tall stature, insulin resistance, and disturbed behavior in a girl with the triple X syndrome harboring three SHOX genes: offspring of a father with mosaic Klinefelter syndrome but with two maternal X chromosomes. *Horm Res*. 2004;61(5):205–210. <http://doi.org/10.1159/000076532>.
30. Lanktree MB, Fantus IG, Hegele RA. Triple X syndrome in a patient with partial lipodystrophy discovered using a high-density oligonucleotide microarray: a case report. *J Med Case Rep*. 2009;3:8867. <http://doi.org/10.4076/1752-1947-3-8867>.
31. van Langevelde K, Flinterman LE, van Hylckama Vlieg A, Rosendaal FR, Cannegieter SC. Broadening the factor V Leiden paradox: pulmonary embolism and deep-vein thrombosis as 2 sides of the spectrum. *Blood*. 2012;120(5):933–946. <http://doi.org/10.1182/blood-2012-02-407551>.