Testicular germ cell tumor susceptibility associated with the \textit{UCK2} locus on chromosome 1q23

Fredrick R. Schumacher$^{1,\dagger}$, Zhaoming Wang$^{2,3,\dagger}$, Rolf I. Skotheim$^{4,5,\dagger}$, Roelof Koster$^{6,\dagger}$, Charles C. Chung$^{2,\dagger}$, Michelle A. T. Hildebrandt$^{8,\dagger}$, Christian P. Kratz$^{9,\dagger}$, Anne C. Bakken$^{4,5}$, D. Timothy Bishop$^{10}$, Michael B. Cook$^{2}$, R. Loren Erickson$^{11}$, Sophie D. Fossa$^{12,\dagger}$, Mark H. Greene$^{2}$, Kevin B. Jacobs$^{2,3}$, Peter A. Kanetsky$^{7}$, Laurence N. Kolonel$^{13}$, Jennifer T. Loud$^{2}$, Larissa A. Korde$^{2,14}$, Loic Le Marchand$^{13}$, Juan Pablo Lewinger$^{1}$, Ragnhild A. Lothe$^{4,5}$, Malcolm C. Pike$^{15}$, Nazneen Rahman$^{16}$, Mark V. Rubertone$^{17}$, Stephen M. Schwartz$^{18}$, Kimberly D. Siegmund$^{1}$, Eila C. Skinner$^{19}$, Clare Turnbull$^{16}$, David J. Van Den Berg$^{1}$, Xifeng Wu$^{6}$, Meredith Yeager$^{2,3}$, Katherine L. Nathanson$^{6}$, Stephen J. Chanock$^{2,\ast}$, Victoria K. Cortessis$^{1,5}$ and Katherine A. McGlynn$^{2,4,\ast}$

$^{1}$Department of Preventive Medicine, Keck School of Medicine, University of Southern California/ Norris Comprehensive Cancer Center, Los Angeles, CA, USA, $^{2}$Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, $^{3}$Cancer Genome Research Laboratory, Division of Cancer Epidemiology and Genetics, Advanced Technology Program, SAIC-Frederick Inc, NCI-Frederick, Frederick, MD, USA, $^{4}$Department of Cancer Prevention, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, $^{5}$Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway, $^{6}$Translational Medicine and Human Genetics, Department of Medicine and $^{7}$Department of Biostatistics and Epidemiology and Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA, $^{8}$Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, $^{9}$Center for Pediatrics and Adolescent Medicine, Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany, $^{10}$Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James’ University Hospital, Leeds, UK, $^{11}$Walter Reed Army Institute of Research, Silver Spring, MD, USA, $^{12}$Departments of Oncology, Oslo University Hospital, The Norwegian Radium Hospital, University of Oslo, Oslo, Norway, $^{13}$Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA, $^{14}$Division of Medical Oncology, University of Washington/Seattle Cancer Care Alliance, Seattle, WA, USA, $^{15}$Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA, $^{16}$Division of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, UK, $^{17}$US Army Center for Health Promotion and Preventive Medicine, Silver Spring, MD, USA, $^{18}$Fred Hutchinson Cancer Research Center and School of Public Health, University of Washington, Seattle, WA, USA and $^{19}$Department of Urology, Stanford University, Stanford, CA, USA

Received October 31, 2012; Revised and Accepted February 27, 2013

Genome-wide association studies (GWASs) have identified multiple common genetic variants associated with an increased risk of testicular germ cell tumors (TGCTs). A previous GWAS reported a possible TGCT susceptibility locus on chromosome 1q23 in the \textit{UCK2} gene, but failed to reach genome-wide significance following replication. We interrogated this region by conducting a meta-analysis of two independent GWASs including a total of 940 TGCT cases and 1559 controls for 122 single-nucleotide polymorphisms

$^\dagger$To whom correspondence should be addressed at: Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. Email: preethi.raj@nih.gov; or chanocks@mail.nih.gov (S.J.C.); Email: mcglynnk@mail.nih.gov (K.A.M).

$^\ddagger$The authors contributed jointly to this work.

$^\ast$These authors jointly directed this work.

Published by Oxford University Press 2013.
(SNPs) on chromosome 1q23 and followed up the most significant SNPs in an additional 2202 TGCT cases and 2386 controls from four case–control studies. We observed genome-wide significant associations for several UCK2 markers, the most significant of which was for rs3790665 ($P_{\text{combined}} = 6.0 \times 10^{-9}$). Additional support is provided from an independent familial study of TGCT where a significant over-transmission for rs3790665 with TGCT risk was observed ($P_{\text{FBAT}} = 2.3 \times 10^{-3}$). Here, we provide substantial evidence for the association between UCK2 genetic variation and TGCT risk.

**INTRODUCTION**

In most developed countries, testicular germ cell tumors (TGCTs) are the most common cancers among men between the ages of 15 and 40 years and incidence rates have been increasing for >50 years (1). The highest rates in the world occur in Norway (12.1/100,000) and Denmark (10.3/100,000) (2). Regardless of the overall incidence in a specific country, however, TGCT incidence is four to five times higher in men of European ancestry than in men of Asian or African ancestry (3). Exogenous risk factors for TGCTs are not yet well elucidated. It is known, however, that risk is increased among men born with undescended testes (4). In addition, men who have had a prior diagnosis of subfertility or TGCT, or who have a family history of TGCT, are at increased risk (5).

The risk of TGCT has been reported to be 8- to 10-fold higher in brothers and 2- to 4-fold higher in sons of men who have had TGCT (6–10). Familial studies have estimated that genetic effects account for nearly a quarter of TGCT risk, which is one of the largest estimated heritabilities reported for any type of cancer (11). Despite the high heritability of TGCT, linkage and candidate gene studies have had limited success identifying TGCT susceptibility loci (12–19). More recently, genome-wide association studies (GWASs) have implicated multiple genomic regions associated with TGCT risk, including those containing KITLG, SPRY4, BAK1, ATF7IP, DMRT1 and TERT (20–23). The discriminative power for TGCT risk using the seven independent GWAS loci plus a rare deletion on the Y chromosome is 69.2% (24), suggesting that additional loci remain undiscovered. Rapley et al. (22) reported a possible TGCT susceptibility locus on chromosome 1. However, the single-nucleotide polymorphism (SNP) rs4657482, which resides within the gene UCK2 on chromosome band 1q23, failed to reach genome-wide significance following replication ($P_{\text{GWAS}} = 1.6 \times 10^{-6}$; $P_{\text{replication}} = 0.07$; $P_{\text{combined}} = 2.0 \times 10^{-6}$). Here, we present the results from a meta-analysis of the UCK2 region from two GWASs of TGCT with additional independent replication that, in turn, have established SNP markers in UCK2 exceeding the threshold for genome-wide significance.

**RESULTS**

To identify susceptibility loci for TGCTs, we conducted a meta-analysis of the GWASs at the National Cancer Institute (NCI) and the University of Southern California (USC). Replication was implemented in studies conducted at the University of Washington (ATLAS study), Oslo University Hospital-Radium Hospital (OUHRH study), MD Anderson Cancer Center (MDA study) and the University of Pennsylvania (TestPAC study) (Table 1 and Supplementary Material, Notes). Further validation of the top UCK2 associations was conducted in a USC TGCT familial study independent of the USC GWAS. In total, the UCK2 meta-analysis included 122 overlapping SNPs in the NCI and USC GWAS among 2499 cases and controls (Table 1). For each of these studies, a 1df trend test for association with TGCT was performed for the 122 UCK2 SNPs assessed in both studies (Supplementary Material, Tables S1 and S2). The combined association tests were generated using a fixed-effects meta-analysis (see the Methods section) and are presented in Supplementary Material, Table S3 for the entire UCK2 region.

In the combined meta-analysis, six SNPs were identified with the corresponding $P$-values $<5.0 \times 10^{-5}$, including the previously reported UCK2 marker rs4657482 ($P_{\text{GWAS}} = 2.25 \times 10^{-7}$; Supplementary Material, Table S3) (22). Five of the top six SNPs were selected for further replication analysis in four additional case–control studies totaling 2202 cases and 2386 controls (Table 2). Two of the studies, the ATLAS study and the TestPAC study, used the iPLEX mass array genotyping platform. The OUHRH study and the MDA study conducted optimized TaqMan genotyping. The pairwise linkage disequilibrium (LD) for the five SNPs among NCI controls ($n = 1056$) showed high LD, except for the SNP rs12562047 (Supplementary Material, Table S4), which is located within an inferred recombination hotspot interval (Fig. 1; chr1:164,090,507-164,097,507). Three of the replication markers (rs12562047, rs4657482 and rs6703280) are located in the first intron of the UCK2 gene. Two of the markers (rs3790665 and rs3790672) are within introns closer to the 3’ region of the gene (Fig. 1), within an interval defined by two recombination peaks identified by five tests of 100 NCI controls without resampling using SequeDhot program (25).

We observed that four of the five tested SNP markers were associated with TGCTs at the level of genome-wide significance ($P < 5.0 \times 10^{-8}$, Table 2). In the combined analysis, the most significant association was observed for rs3790665 ($P_{\text{combined}} = 6.0 \times 10^{-9}$) with a summary odds ratio (OR) estimate of 1.26 (95% confidence interval, CI, 1.17–1.37) per copy of the C allele (Supplementary Material, Fig. S1). The effects were similar across all of the studies ($P_{\text{het}} = 0.03$), except for one study where a null signal was observed. By examining these results in an independent family study, we observed a significant over-transmission ($P_{\text{FBAT}} = 2.3 \times 10^{-3}$) to the TGCT cases for the rs3790665 risk allele, C (Supplementary Material, Table S5).

The second most significant finding from the combined analysis was for SNP rs4657482 ($P_{\text{combined}} = 1.2 \times 10^{-8}$), the...
marker originally reported by Rapley et al. (22). The combined OR estimate from the meta-analysis for this SNP was 1.25 (95% CI 1.16–1.35) per copy of the A allele. The corresponding estimates reported by Rapley et al. were 1.39 and 1.14 for their GWAS and replication studies, respectively (22). We also observed a significant over-transmission of the A allele to TGCT cases in the family study ($P_{\text{Fisher}} = 5.9 \times 10^{-3}$; Supplementary Material, Table S5). The associations reported by Rapley et al. (22) are independent of the associations presented here. A meta-analysis of rs4657482 across our combined association analysis and the combined association reported by Rapley et al. (22) yielded a highly significant association $[\text{OR} = 1.26 \ (95\% \ CI \ 1.18–1.33); \ P = 1.3 \times 10^{-13}]$. When we examined the set of five markers sequentially in conditional analyses, we did not observe any clear evidence for a second independent signal (Supplementary Material, Table S6). Although there is a weak correlation between rs12562047 and rs3790665 ($r^2 = 0.18$), the former maps within an inferred hotspot and the conditional analysis does not support a second signal. Haplotype analysis for these five markers using HapMap (v3, release 27) CEU/TSI data estimates the haplotype frequency in which the effect allele (G) for rs12562047 resides with two or more risk alleles was 30.5%. Further sequence and fine-mapping analysis is required to define the underlying haplotype, as each of these SNPs reported herein most likely represent markers and not the directly associated variant.

In Supplementary Material, Table S7, we report the combined meta-analysis results for the known TGCT susceptibility loci in the NCI and USC GWAS. Single or multiple markers reached genome-wide significance for all of the known regions, except the ATF7IP gene region ($P = 0.065$). Of the significant loci, markers in the KITLG region were the most significant ($P = 1.25 \times 10^{-23}$) followed by the SPRY4 locus ($P = 6.4 \times 10^{-11}$). Rapley et al. reported a borderline TGCT susceptibility locus on chromosome 4 for rs4699052. Although our combined meta-analysis failed to reach genome-wide significance for this marker ($P = 3.5 \times 10^{-3}$), further evaluation is needed.

### DISCUSSION

In a multi-stage analysis of SNPs in the UCK2 region of chromosome 1, we confirm a TGCT susceptibility locus at 1q23, which harbors the plausible candidate gene uridine/cytidine kinase-2. A marker in this region was previously implicated in TGCT but failed to reach genome-wide significance (22). We provide here further support for this locus by confirming associations of the most significantly associated variants in an independent family-based study. The most promising SNP, rs3790665, lies in the intronic region of UCK2 near the 3’ end of the gene (Fig. 1). UCK2, initially identified as a testis-specific gene TSA903 (26), encodes a pyrimidine ribonucleoside kinase that catalyzes the phosphorylation of uridine and cytidine to form uridine monophosphate and cytidine monophosphate (27). UCK2 contains seven exons and spans a 19 kb region on cytoband 1q23 and codes for a 261-amino acid protein. Other reported GWASs have not implicated UCK2 with non-TGCT traits.

Using the ENCODE resources (28), including HaploReg (29) and RegulomeDB (30) (Supplementary Material, Table S8), we evaluated 26 surrogate SNP markers in high LD with the five markers reported here ($r^2 > 0.8$, max distance = 200 kb, 1000 Genomes CEU). Interestingly, five SNPs directly map to sequences reported to be within a region enriched for enhancers, noted across a spectrum of tested cell types (Supplementary Material, Table S8). In addition, rs10918304 ($r^2 = 0.88$ with rs3790665; $r^2 = 0.82$ with

---

**Table 1.** Total number of TGCT cases and controls included in the meta- and replication analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Total number</th>
<th>Controls</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS meta-analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCI</td>
<td>1638</td>
<td>1056</td>
<td>582</td>
</tr>
<tr>
<td>USC</td>
<td>861</td>
<td>503</td>
<td>358</td>
</tr>
<tr>
<td>Total GWAS</td>
<td>2499</td>
<td>1559</td>
<td>940</td>
</tr>
<tr>
<td>Replication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATLAS</td>
<td>1399</td>
<td>960</td>
<td>439</td>
</tr>
<tr>
<td>OUHRR</td>
<td>1257</td>
<td>392</td>
<td>865</td>
</tr>
<tr>
<td>MDA</td>
<td>590</td>
<td>351</td>
<td>239</td>
</tr>
<tr>
<td>TestPAC</td>
<td>1342</td>
<td>683</td>
<td>659</td>
</tr>
<tr>
<td>Total replication</td>
<td>4588</td>
<td>2386</td>
<td>2202</td>
</tr>
<tr>
<td>Total all</td>
<td>7087</td>
<td>3945</td>
<td>3142</td>
</tr>
</tbody>
</table>

**Table 2.** Meta-analysis and replication results for UCK2 variants

<table>
<thead>
<tr>
<th>SNP (EA)</th>
<th>Chr:position</th>
<th>Stage</th>
<th>OR (95% CI)</th>
<th>Stage</th>
<th>$P$-value</th>
<th>Combined</th>
<th>Heterogeneity ($\phi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12562047-G</td>
<td>1:164095780</td>
<td>GWAS</td>
<td>0.77 (0.68–0.87)</td>
<td>4.2 $\times 10^{-5}$</td>
<td>1.1 $\times 10^{-4}$</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>rs4657482-A</td>
<td>1:164098273</td>
<td>Replication</td>
<td>0.92 (0.84–1.01)</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6703280-T</td>
<td>1:164107035</td>
<td>Replication</td>
<td>1.39 (1.23–1.57)</td>
<td>1.2 $\times 10^{-7}$</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3790665-C</td>
<td>1:164136695</td>
<td>GWAS</td>
<td>1.37 (1.20–1.56)</td>
<td>3.6 $\times 10^{-6}$</td>
<td>6.0 $\times 10^{-9}$</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>rs3790672-C</td>
<td>1:164140016</td>
<td>Replication</td>
<td>1.35 (1.19–1.54)</td>
<td>1.7 $\times 10^{-8}$</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Recombination plot and linkage disequilibrium structure for the TGCT susceptibility region at the UCK2 locus. Regional plot of association results, recombination hotspots and linkage disequilibrium for the UCK2 locus. TGCT susceptibility region. Combined meta-analysis results are shown as red diamonds with rsID labeled, replication result in royal blue and NCI-USC GWAS meta-analysis in gray. For association results, $-\log_{10}P$-values ($y$-axis, left) of the SNPs are shown according to their chromosomal positions ($x$-axis). Linkage disequilibrium structure based on NCI controls ($n = 1,188$) was visualized by snp.plotter software. The line graph shows likelihood ratio statistics ($y$-axis, right) for recombination hotspot by SequenceLDhot software and five different colors represent five tests of 100 controls from NCI without resampling. Physical locations of each region are based on NCBI Build 36 of the human genome.
had to fulfill the following to be included: SNP call rate >95%, no deviation from Hardy–Weinberg equilibrium in controls at \( P < 0.00001 \); <2% discordance between genotypes in duplicate. Cluster plots for SNPs that were close to failing any of the QC criteria were re-examined centrally.

Statistical methods

Each GWAS assessed genetic associations using a 1df trend test assuming an additive genetic inheritance model adjusting for population ancestry, age and study using PLINK for the USC GWAS and GLU (Genotyping Library and Utilities; http://code.google.com/p/glu-genetics/) for the NCI GWAS. ORs and 95% CIs were estimated using unconditional logistic regression. Replication case–control studies were analyzed using unconditional logistic regression. A fixed-effect meta-analysis was performed weighting each study contribution by the inverse of the standard error. A conservative genome-wide level of statistical significance (\( P < 5.0 \times 10^{-8} \)) was applied for the combined association analysis to minimize the chance of false-positive results. Within the USC TGCT familial study, a family-based association was tested comparing the distribution of alleles transmitted to affected individuals with the distribution of those alleles that are not transmitted. The over-transmission and under-transmission of each locus were assessed for the replication SNPs. The analytical software FBAT (33) was used for the familial testing.

MATERIALS AND METHODS

Samples

The discovery meta-analysis for the UCK2 region (Chr 1: 163,801,412-164,192,158) was conducted using two GWAS datasets, NCI and USC (Table 1). In total 122 SNPs overlapped the two GWAS spanning the UCK2 region on chromosome 1. Analyses were based on the datasets following standard quality control procedures (Supplementary Material, Notes). We chose five of the six most significant SNPs in our UCK2 meta-analysis for independent replication. The replication stage included 2202 cases and 2386 controls from four TGCT case–control studies (Table 1 and Supplementary Material, Notes). An independent set of TGCT pedigrees had been enrolled at USC by identifying TGCT probands and recruitment of eligible family members. All studies were limited to non-Hispanic whites and approved by the appropriate ethics committees.

Genotyping

Two replication studies (OUHRH and MDA) and the family studies were genotyped using the 5′-exonuclease assay (Taqman\textsuperscript{TM}) and the ABI Prism 7900HT sequence detection system, all according to the manufacturer’s instructions, across several genotyping centers. Primers and probes were supplied directly by Applied Biosystems as Assays-By-Design\textsuperscript{TM}. The ATLAS and TestPAC studies conducted genotyping using the iPLEX mass array platform. Assays at all genotyping centers included at least four negative controls and 2–5% duplicates on each 384-well plate. Data on each SNP for a given site

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

Conflict of Interest statement. None declared.

FUNDING

A portion of this work was supported by the Intramural Research Program of the National Cancer Institute and by support services contract HHSN261200655004C with Westat, Inc. Support was also provided as follows: USC GWAS controls were supported by the Multiethnic Cohort Study (NCI U01-CA98758). USC GWAS testicular cases and Familial Study were supported by the California Cancer Research Program (99-00505V-10260, 03-00174VRS-30021) and National Cancer Institute (1R01 CA102042-01A1) grants to V.K.C. and a Whittier Foundation award to the Norris Comprehensive Cancer Center. Replication effort for the TestPAC study was supported by the Abramson Cancer Center at the University of Pennsylvania and National Institutes of Health grant R01CA114478 to P.A.K. and K.L.N. Replication effort for the ATLAS study was supported by the National Institutes of Health grant R01CA085914 to S.M.S. MD Anderson: Center for Translational and Public Health Genomics of the Duncan Family Institute for Cancer Prevention and Risk Assessment and by MD Anderson Senior Research Trust...
Fellowship to X.W. The UK Genetics of Testicular Cancer Study was supported by Movember and Cancer Research UK.

REFERENCES


