



Published in final edited form as:

*Eur Urol.* 2012 February ; 61(2): 258–268. doi:10.1016/j.eururo.2011.10.007.

## Meta-analysis of Clear Cell Renal Cell Carcinoma Gene Expression Defines a Variant Subgroup and Identifies Gender Influences on Tumor Biology

A. Rose Brannon<sup>a,b</sup>, Scott M. Haake<sup>a,c</sup>, Kathryn E. Hacker<sup>a,b</sup>, Raj S. Pruthi<sup>a,d</sup>, Eric M. Wallen<sup>a,d</sup>, Matthew E. Nielsen<sup>a,d</sup>, and W. Kimryn Rathmell<sup>a,b,c,\*</sup>

<sup>a</sup>Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

<sup>b</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC, USA

<sup>c</sup>Department of Medicine, University of North Carolina, Chapel Hill, NC, USA

<sup>d</sup>Division of Urologic Surgery, University of North Carolina, Chapel Hill, NC, USA

### Abstract

**Background**—Clear cell renal cell carcinoma (ccRCC) displays molecular and histologic heterogeneity. Previously described subsets of this disease, ccA and ccB, were defined based on multigene expression profiles, but it is unclear whether these subgroupings reflect the full spectrum of disease or how these molecular subtypes relate to histologic descriptions or gender.

**Objective**—Determine whether additional subtypes of ccRCC exist and whether these subtypes are related to von Hippel-Lindau (VHL) inactivation, hypoxia-inducible factor (HIF) 1 and 2 expression, tumor histology, or gender.

**Design, setting, and participants**—Six large, publicly available ccRCC gene expression databases were identified that cumulatively provided data for 480 tumors for meta-analysis via meta-array compilation.

---

© 2011 European Association of Urology. Published by Elsevier B.V. All rights reserved.

\*Corresponding author. 450 West Dr, CB 7295, Lineberger Comprehensive Cancer Center, Chapel Hill, NC 27599 USA. Tel: 919-966-3522; Fax: 919-966-8212. rathmell@med.unc.edu (W.K. Rathmell).

**Author contributions:** W. Kimryn Rathmell had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Brannon, Hacker, Haake, Rathmell.

**Acquisition of data:** Brannon, Hacker, Haake, Pruthi, Wallen, Nielsen.

**Analysis and interpretation of data:** Brannon, Hacker, Haake, Rathmell.

**Drafting of the manuscript:** Brannon, Hacker, Haake, Rathmell.

**Critical revision of the manuscript for important intellectual content:** Brannon, Hacker, Haake, Pruthi, Wallen, Nielsen, Rathmell.

**Statistical analysis:** Brannon, Hacker, Haake.

**Obtaining funding:** Rathmell.

**Administrative, technical, or material support:** Pruthi, Wallen, Nielsen.

**Supervision:** Rathmell.

**Other (specify):** None.

**Financial disclosures:** I certify that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Measurements**—Unsupervised consensus clustering was performed on the meta-arrays. Tumors were examined for the relationship of multigene-defined consensus subtypes and expression signatures of VHL mutation and HIF status, tumor histology, and gender.

**Results and limitations**—Two dominant subsets of ccRCC were observed. However, a minor third cluster was revealed that correlated strongly with a wild type (WT) VHL expression profile and indications of variant histologies. When variant histologies were removed, ccA tumors naturally divided by gender. This technique is limited by the potential for persistent batch effect, tumor sampling bias, and restrictions of annotated information.

**Conclusions**—The ccA and ccB subsets of ccRCC are robust in meta-analysis among histologically conventional ccRCC tumors. A third group of tumors was identified that may represent a new variant of ccRCC. Within definitively clear cell tumors, gender may delineate tumors in such a way that it could have implications regarding current treatments and future drug development.

### Keywords

Clear cell renal cell carcinoma; Gene expression; Gender; Hypoxia; HIF; Renal cell carcinoma; RCC; VHL

## 1. Introduction

More than 60% of kidney tumors are histologically diagnosed as clear cell renal cell carcinoma (ccRCC) [1]. It has become increasingly apparent that ccRCC displays significant heterogeneity at the molecular, histologic, and clinical levels [2]. We and others have previously demonstrated that ccRCC is composed of at least two subtypes, with different patterns of gene expression and different clinical outcomes [3,4]. Our group used biologically driven clustering to define two robust subgroups of ccRCC, ccA and ccB, that are highly dichotomous by molecular phenotype and cancer-specific survival (8.6 yr vs 2 yr, respectively;  $p = 0.003$ ) [5]. Other studies have identified transcript patterns related to expression of hypoxia inducible factors (HIF) 1 and 2, regulated by the von Hippel Lindau (VHL) tumor suppressor [6,7], metastatic characteristics [8], and genetic sequence [9].

Gender, however, has not been explored as a factor influencing tumor biology, despite the well-known gender disparity in this disease [10]. ccRCC tumors in men display more aggressive features than in women [11,12], progress to metastatic disease after nephrectomy twice as frequently [13], and have decreased tumor-specific and overall survival [10,12,14]. The overrepresentation and poor prognosis for men provide clues that the disease may differ between genders.

To determine the breadth of ccRCC subtypes that may be relevant for genetic discoveries or clinical outcome predictions, it is necessary to compile larger datasets for meta-analyses. We therefore generated datasets, or *meta-arrays*, of available gene expression studies involving 480 tumors. Tumors with variant clear cell histology segregate separately from ccA or ccB and display a genetic program consistent with maintenance of a wild type (WT) VHL. These two subsets remain the most robust subdivisions of ccRCC. Gender is also strongly related with the genomic characteristics of tumors, shedding light on important tumor features differentially presented in males and females.

## 2. Methods

### 2.1. Study selection

A literature review was performed for ccRCC gene expression analyses, yielding 29 published studies. Studies were excluded for lack of clinical data, fewer than 20 tumors, redundant analysis of previously published data, or fewer than 5000 genes analyzed. Six studies were included in the final analysis [4–9] (Table 1). Approval by institutional ethics review panels was documented in each manuscript.

### 2.2. Additional sample preparation

Forty-four additional ccRCC quality-checked, flash-frozen nephrectomy specimens from the University of North Carolina (UNC) were accessed with approval from the UNC Biomedical Institutional Review Board. RNA was prepared as previously described [5] and hybridized against a common reference (Perou et al [15]) on Agilent Whole Human Genome (4 × 44 K) microarrays (Agilent Technologies, Santa Clara, CA, USA). Data were uploaded to the UNC Microarray Database (UMD; <https://genome.unc.edu>).

### 2.3. Data preprocessing

Data from tumors previously analyzed by Brannon et al [5] were redownloaded with the 44 new samples from the UMD in log<sub>2</sub> Lowess-normalized sample/reference format (median). Entrez gene ID was reannotated from the 20101031 Agilent annotation release.

Sample data from Zhao et al [4] were redownloaded from the Stanford Microarray Database (<http://smd.stanford.edu>) with Entrez ID annotation as log<sub>2</sub> normalized ratios (median). Raw data from the remaining studies used Affymetrix arrays and were retrieved from the National Center for Biotechnology Information Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo>). For the Wuttig et al dataset, metastatic tumors were excluded. Raw data files were imported into Partek Genomics Suite v.6.5 software (Partek, St. Louis, MO, USA) using RMA normalization. Each Affymetrix array type was imported separately, and data were subsequently merged into one file containing only the overlapping probes. The outlier effect of each sample was analyzed by principal components. Three samples (*VARI\_038T*, *VARI\_046T*, and *VARI111T*) were removed from the Dalglish et al [9] data for being extreme outliers. Entrez ID annotation was performed using the Affymetrix annotation release 31 (HG-U133 plus 2). Individual datasets were filtered for 70% probes above background levels and medians of redundant Entrez IDs calculated.

### 2.4. Compilation of tumor meta-arrays

Entrez IDs common to all datasets were identified using MergeMaid in R, and the Affymetrix data were merged with the Agilent datasets. This combined dataset was imported into the Partek Genomics Suite and the dataset source annotated for each sample. Series 1 and series 2 in Dalglish et al [9] were processed on different chip series at different sites and thus were marked as different sources. Once annotated, batch effects for each dataset source were removed simultaneously through Partek's batch effect removal algorithms. Principal components analysis (PCA) plots were created using the Partek Genomics software (Fig. 1).

All tumors were used to create the T480 array, composed of 6386 overlapping genes. Tumors with ambiguous histology (on re-review or based on annotation not specifically ccRCC or labeled ccRCC in one location but non-clear cell in another) were removed for the building of the T418 meta-array composed of 6394 genes and meta-array T261 (to maximize genome coverage) with 11–742 genes.

## 2.5. Compilation of the normal meta-array

To expand the numbers of normal samples for comparison, two gender-annotated normal kidney series (GSE781 and GSE11024) were combined with studies used in the tumor meta-array that included normal tissue, excluding tissues from individuals <18 yr of age. Array compilation used methods as described for the tumor meta-array.

## 2.6. Tumor assignment

The National Cancer Institute (NCI) Biometric Research Branch (BRB) ArrayTools v.4.2.0 beta 1 software (<http://linus.nci.nih.gov/BRB-ArrayTools.html>) was used to individually assign tumors in the T480 array as ccA, ccB, or unclassified based on 77 present reannotated (Agilent version 20101031) Entrez IDs from the previously identified biomarker probes [5]. Cut points were determined using the previously assigned tumors to account for changes from batch normalization and genes univariately significant at  $p < 0.001$ . Tumors were assigned to a subtype if all seven class prediction algorithms yielded the same subtype call.

HIF1 and HIF2, HIF2 only, or WT assignment was made in BRB ArrayTools starting with all available genes and using the samples from Gordan et al [6] for cut point creation. Tumors were assigned based on the consensus of five of seven calls from the class-prediction algorithms at  $p < 0.001$ , prediction analysis for microarrays (PAM), and binary tree prediction, allowing assignment of 370 of the 480 tumors.

## 2.7. Identification of clusters

We used ConsensusCluster [16] for PCA before consensus clustering the data. PCA eigenvectors representing 85% of the variation in data were calculated, and genes with coefficients in the top 25% were chosen for clustering. Consensus clustering was performed using k-means and 80% bootstrapping genes or samples with >500 iterations.

## 2.8. Data analyses

Genes (false discovery rate <0.00001) and biologic pathways ( $p < 0.05$ ) underscoring clustering patterns were analyzed using Significance Analysis of Microarrays (SAM; <http://www-stat.stanford.edu/~tibs/SAM>), SAM gene set analysis (GSA), Gene Set Enrichment Analysis (GSEA; <http://www.broad.mit.edu/gsea>), single-sample GSEA [17], and gene sets from the Broad Institute (<http://www.broad.mit.edu/gsea/msigdb>). SAM was run with 1000 permutations. SAM-GSA and GSEA were run with 5000 permutations. For gender analysis, SAM was run with all genes; SAM-GSA was run with only autosomal genes. Genes were annotated with location and function using Source (<http://source.stanford.edu>) or the GeneCards Human Gene Database (<http://genecards.org>). The gender-related genes underwent functional annotation clustering in the Database for Annotation, Visualization, and Integrated Discovery (DAVID) v.6.7 (<http://david.abcc.ncifcrf.gov>). Heat maps were produced using Cluster v.3.0 (<http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster>) and Java TreeView (<http://jtreeview.sourceforge.net>).

## 3. Results

### 3.1. Identification of a third subcluster with a wild type von Hippel-Lindau signature

To address whether additional subtypes of ccRCC exist beyond ccA and ccB, we analyzed a meta-array of 480 ccRCC samples encompassing 6386 genes [4–9]. Even with 480 tumors, ccRCC remained dominated by the two major subtypes (Fig. 1c), corresponding in gene signature profile to ccA and ccB. Adjusting the subclustering algorithm to identify

additional groups, a third cluster (Cluster3) comprising 67 ccRCC tumors is present (Fig. 1d). No additional clusters were present (Fig. 1e–g).

Cluster3 tumors overexpress many gene sets, such as those related to estrogen-related receptor alpha (*ESRRα*) targets and mitochondrial bioenergetics (Fig. 2a) [18] (Supplemental Table 1). Compared to ccA and ccB tumors, Cluster3 tumors underexpress hypoxia and angiogenesis gene sets. Expression profiles indicative of WT VHL (*WT VHL* gene), HIF1 and HIF2 overexpression (*H1H2*), or HIF2 only (*H2*) overexpression [6] were used to classify the tumors. The larger subtypes are not divided based on expression defined VHL or HIF profiles but 82% of Cluster3 classified as WT VHL (Fig. 2).

### 3.2. Cluster3, not clearly clear cell

This predominance of tumors with a WT VHL signature led us to question whether these tumors represented histologic classification ambiguity. Of samples processed at UNC with slides available (71 tumors), eight were classified in Cluster3, and six of these displayed deviations from classic clear cell features (Fig. 3), typically demonstrating papillary architecture with clear cells. All were diagnosed as clear cell in the pathology reports, although the term *papillary architecture* was included in the histologic description in one case. In contrast, variant histology was only identified in one of the remaining 63 tumor sections. When the remaining arrays were cross-referenced to all sources of reported histologic data, more than half of the Cluster3 set was reported ambiguously. These results suggest that a substantial subset of tumors nominally classified as ccRCC may have an ambiguous non-clear cell phenotype and may represent a distinct subclassification for which current histologic evaluation alone is inadequate. However, gene expression analysis is able to confidently detect this group.

### 3.3. Only ccA and ccB remain in definitively clear cell arrays

To determine whether additional subtypes remained in the absence of any tumors with ambiguous histologic designations (62 tumors removed), two new meta-arrays were compiled (T418 with 418 samples and 6394 genes and T261 with 261 samples using only more recently constructed arrays to maximize genomic coverage to 11–742 genes).

Only two clusters, ccA and ccB, were present even when additional clusters were forced (Fig. 4). The disappearance of the third subcluster upon removal of ambiguous histologies strongly suggests that ccA and ccB are inherently stable and likely highly divergent.

### 3.4. ccA tumors naturally divide by gender

We then sought to determine whether additional structure existed within the subtypes. When unsupervised consensus clustering was applied to the T261 ccA subset tumors (Fig. 5), two distinct subgroups emerged along gender lines, with one cluster 98% (81 of 83) male and the other 85% (63 of 74) female. A similar pattern was observed in T418 (data not shown). These observations led us to surmise that strong transcriptomic signals are associated with gender in ccRCC.

### 3.5. Tumors from males and females collectively present different gene ontologies, particularly immunologic surveillance and catabolic processes

The full meta-arrays were then analyzed based on gender, identifying 326 differentially expressed autosomal genes (366 total genes; Fig. 6a). Several genes identified are potentially important to tumor biologic differences observed between genders (Table 2), including genes involved in DNA repair (*PARP8*, *ERCC5*), cell signaling (*MAPKSP1*, *SPAG1*, *Fyn*, *Fos*, and others), and motility (*VCAN*, *CYR61*).

GSEAs in tumors from males demonstrated overexpression of many immune or inflammatory gene sets (Table 3). These results may reflect tumor infiltration by lymphocytes or another immune cell subset. In contrast, females' tumors overexpressed catabolic process-related genes, indicating a potentially distinct metabolic pattern (Table 3). The complete lists of genes differentially expressed in the T261 and T418 meta-arrays are provided in Supplemental Tables 2 and 3, respectively.

Finally, concerned about the inclusion of intrinsic gender differences in our tumor analyses, we analyzed 52 normal kidneys. Few genes were differentially regulated between the genders in normal tissues and were almost exclusively associated with sex chromosome loci, with no overlap with the genes or gene sets reported in the tumor analysis (Supplemental Table 4).

#### 4. Discussion

The compilation of multiple datasets provides a unique opportunity to explore transcript variation in ccRCC. Significant to the field and to clinicians managing the treatment of these cancers, these data confirm ccRCC division into two dominant groups: ccA and ccB. Further, a variant subgroup also exists that likely represents a highly divergent group, characterized by WT VHL status. This type of combinatorial computational analysis must be considered within the context of numerous inherent constraints. First, gene expression distinctions are dependent on the information attained from a small portion of the tumor and may not correlate with differences in tumor biology, cancer behavior, or actionable biomarkers, so that these findings require independent mechanistic validation. Although the PCA following batch effect removal demonstrated a homogenous population, these samples undergo different protocols of collection, storage, and processing. In addition, the platforms on which samples were analyzed are distinct, as are the gene sets, which may lead to residual batch effects. Also, tissue samples were not widely available, so that rigorous pathologic re-review of histology was not possible for the majority of tumors.

That genders have different physiologic and hormonal environments that could affect carcinogenesis is not surprising, but our study suggests that host immune surveillance may also be relevant to tumor biology. This finding is of particular importance given the role of the immune system (and its stimulation) in the surveillance and treatment of ccRCC and the implications of inflammatory cells within tumors [19,20]. Moreover, gender-specific differences in immune regulation have long been appreciated and recently extended to pathways currently under investigation as targets for ccRCC treatment [21]. Alternatively, these gene sets may represent gender-specific immune phenomena not well understood at this juncture. In contrast, female-derived tumors demonstrate a stronger catabolic fingerprint, suggesting that female-derived tumors may be more closely aligned with mechanisms that promote a metabolic switch. In addition, metabolic properties of tumors are gaining traction as an important way to potentially target cells therapeutically [22].

This analysis identified a small but distinct subset of tumors, along with evidence that histologic clues were present, indicating these tumors were not the same as conventional ccRCC and that they possess WT VHL. These observations suggest that ambiguous histology is not a sufficient surrogate for ccRCC histologic diagnosis. Further efforts to refine pathologic conventions for renal cancer diagnosis will be an important future consideration. Given that Cluster3 tumors share the WT VHL feature, ccRCC should perhaps be redefined as a tumor with unambiguous clear cell histology or evidence of VHL gene inactivation. Ambiguous or hybrid tumors require independent consideration for therapy selection, much as non-clear cell histology tumors are currently managed. These results indicate a need for expert assessment of tumor histology in making a diagnosis of

ccRCC and have important implications for clinical trial eligibility criteria for ccRCC studies. Ultimately, the incorporation of molecular testing into routine pathologic diagnosis will clarify these ambiguities to the benefit of patient care.

Intriguingly, Cluster3 tumors overexpress ESRR $\alpha$ -regulated transcripts and gene sets involved in mitochondrial metabolism. ESRR $\alpha$  is gaining recognition in the cancer community as an important mediator of tumor metabolism, driving transcriptional activation of many genes involved in maintaining mitochondrial biogenesis and oxidative phosphorylation [23,24]. As inhibitors of this signaling pathway are rapidly emerging, it is important to consider that even this smaller subset of VHL WT RCC tumors might benefit from greater understanding of their tumor biology.

## 5. Conclusions

Overall, this meta-analysis demonstrates that diagnosis of ccRCC requires expert pathology input and that WT VHL signatures correlate with a subgroup that is simultaneously transcriptionally unique and histologically divergent as well as displaying a distinct profile of metabolic genes. Conventional ccRCC remains dominated by the presence of two robust subgroups, corresponding to the previously described ccA and ccB. These data provide valuable confirmation in a much larger set of tumors than ever before explored that ccA and ccB are the major source of tumor division, with gender disparity providing additional disease information. These findings will have important implications for personalized disease management.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

*Funding/Support and role of the sponsor:* Drs. Rathmell and Brannon received funding from the American Association for Cancer Research Landon INNOVATOR Award; Dr. Hacker received funding from the National Institutes of Health Medical Scientist Training Program, T32 GM008719; and Dr. Rathmell received funding from NCI, R01 CA121781.

The authors thank Dr. Pablo Tamayo, Broad Institute, for single-sample GSEA R scripts; Dr. Signoretti, Dana Farber Cancer Institute, for providing post-publication specimen annotations; the UNC Lineberger Tissue Procurement Facility; the UNC Genomics and Microarray analysis core; the UNC Tissue Pathology Laboratory; and the UNC Lineberger Cancer Center for support of these essential core facilities.

## References

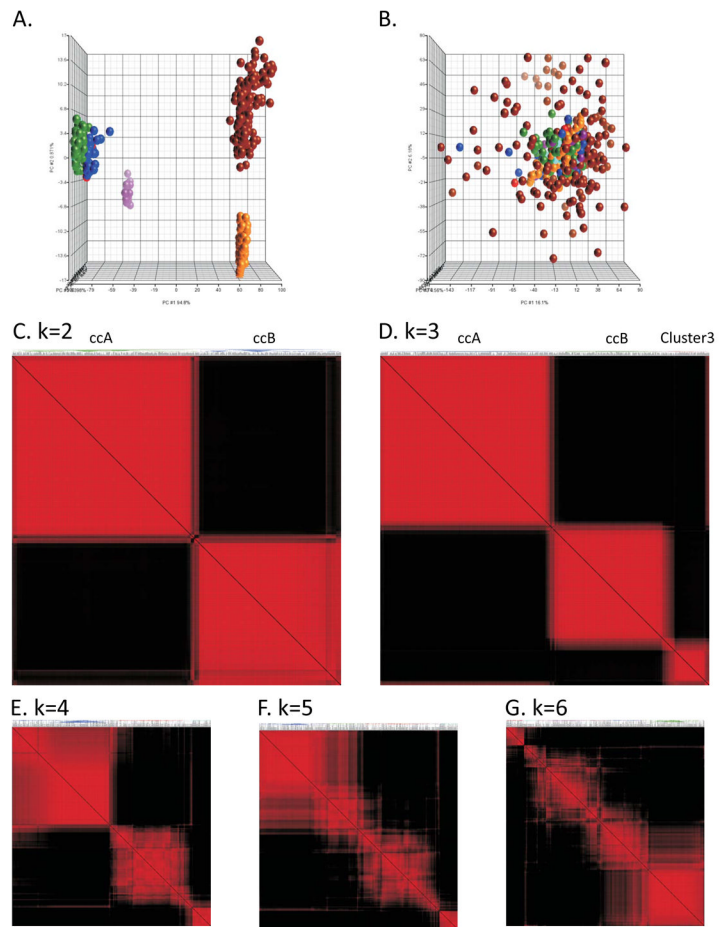
1. Palapattu GS, Pantuck AJ, Dorey F, Said JW, Figlin RA, Belldegrun AS. Collecting system invasion in renal cell carcinoma: impact on prognosis and future staging strategies. *J Urol.* 2003; 170:768–72. discussion 772. [PubMed: 12913694]
2. Oosterwijk E, Rathmell WK, Junker K, et al. Basic research in kidney cancer. *Eur Urol.* 2011; 60:622–33. [PubMed: 21741760]
3. Brannon AR, Rathmell WK. Renal cell carcinoma: where will the state-of-the-art lead us? *Curr Oncol Rep.* 2010; 12:193–201. [PubMed: 20425079]
4. Zhao H, Ljungberg B, Grankvist K, Rasmuson T, Tibshirani R, Brooks JD. Gene expression profiling predicts survival in conventional renal cell carcinoma. *PLoS Med.* 2006; 3:e13. [PubMed: 16318415]
5. Brannon AR, Reddy A, Seiler M, et al. Molecular stratification of clear cell renal cell carcinoma by consensus clustering reveals distinct subtypes and survival patterns. *Genes Cancer.* 2010; 1:152–63. [PubMed: 20871783]

6. Gordan JD, Lal P, Dondeti VR, et al. HIF-alpha effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma. *Cancer Cell*. 2008; 14:435–46. [PubMed: 19061835]
7. Beroukhim R, Brunet JP, Di Napoli A, et al. Patterns of gene expression and copy-number alterations in von-Hippel Lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Res*. 2009; 69:4674–81. [PubMed: 19470766]
8. Wuttig D, Baier B, Fuessel S, et al. Gene signatures of pulmonary metastases of renal cell carcinoma reflect the disease-free interval and the number of metastases per patient. *Int J Cancer*. 2009; 125:474–82. [PubMed: 19391132]
9. Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*. 2010; 463:360–3. [PubMed: 20054297]
10. Stafford HS, Saltzstein SL, Shimasaki S, Sanders C, Downs TM, Sadler GR. Racial/ethnic and gender disparities in renal cell carcinoma incidence and survival. *J Urol*. 2008; 179:1704–8. [PubMed: 18343443]
11. Chen J, Shi B, Zhang D, Jiang X, Xu Z. The clinical characteristics of renal cell carcinoma in female patients. *Int J Urol*. 2009; 16:554–7. [PubMed: 19456990]
12. Schrader AJ, Sevinc S, Olbert PJ, Hegele A, Varga Z, Hofmann R. Gender-specific characteristics and survival of renal cell carcinoma [in German]. *Urologe A*. 2008; 47:1182, 1184–6. [PubMed: 18679648]
13. Abel EJ, Culp SH, Meissner M, Matin SF, Tamboli P, Wood CG. Identifying the risk of disease progression after surgery for localized renal cell carcinoma. *BJU Int*. 2010; 106:1277–83. [PubMed: 20394619]
14. Onishi T, Oishi Y, Goto H, Yanada S, Abe K. Gender as a prognostic factor in patients with renal cell carcinoma. *BJU Int*. 2002; 90:32–6. [PubMed: 12081765]
15. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406:747–52. [PubMed: 10963602]
16. Seiler M, Huang CC, Szalma S, Bhanot G. ConsensusCluster: a software tool for unsupervised cluster discovery in numerical data. *OMICS*. 2010; 14:109–13. [PubMed: 20141333]
17. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009; 462:108–12. [PubMed: 19847166]
18. Scarpulla RC. Transcriptional paradigms in mammalian mitochondrial biogenesis and function. *Physiol Rev*. 2008; 88:611–38. [PubMed: 18391175]
19. Bleumer I, Oosterwijk E, De Mulder P, Mulders PF. Immunotherapy for renal cell carcinoma. *Eur Urol*. 2003; 44:65–75. [PubMed: 12814677]
20. Coppin C, Porzolt F, Awa A, Kumpf J, Coldman A, Wilt T. Immunotherapy for advanced renal cell cancer. *Cochrane Database Syst Rev*. 2005:CD001425. [PubMed: 15674877]
21. Dinesh RK, Hahn BH, Singh RP. PD-1, gender, and autoimmunity. *Autoimmun Rev*. 2010; 9:583–7. [PubMed: 20433954]
22. Kroemer G, Pouyssegur J. Tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell*. 2008; 13:472–82. [PubMed: 18538731]
23. Mirebeau-Prunier D, Le Pennec S, Jacques C, et al. Estrogen-related receptor alpha and PGC-1-related coactivator constitute a novel complex mediating the biogenesis of functional mitochondria. *FEBS J*. 2010; 277:713–25. [PubMed: 20067526]
24. Wang J, Wang Y, Wong C. Oestrogen-related receptor alpha inverse agonist XCT-790 arrests A549 lung cancer cell population growth by inducing mitochondrial reactive oxygen species production. *Cell Prolif*. 2010; 43:103–13. [PubMed: 20447055]

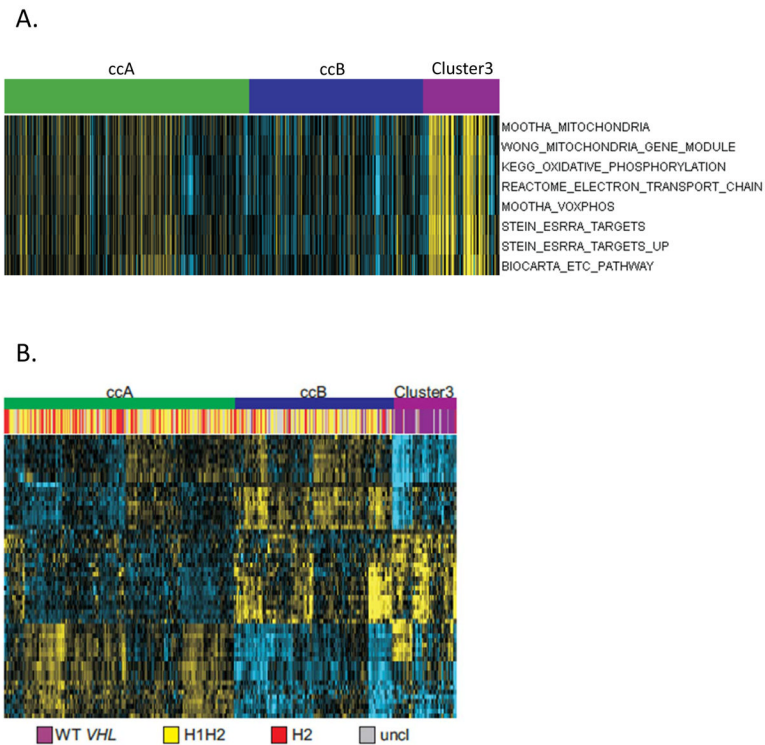


**Take-home message**

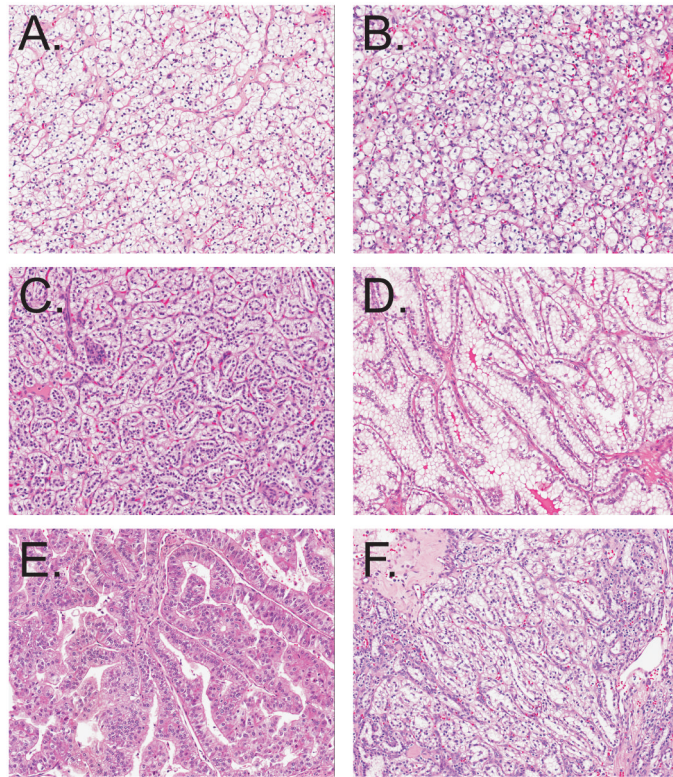
Clear cell renal cell carcinoma comprises two dominant subtypes, ccA and ccB, with gender disparity providing additional disease information. A third minor subgroup has distinct expression profiles consistent with von Hippel-Lindau wild type status and displays variant histology features.



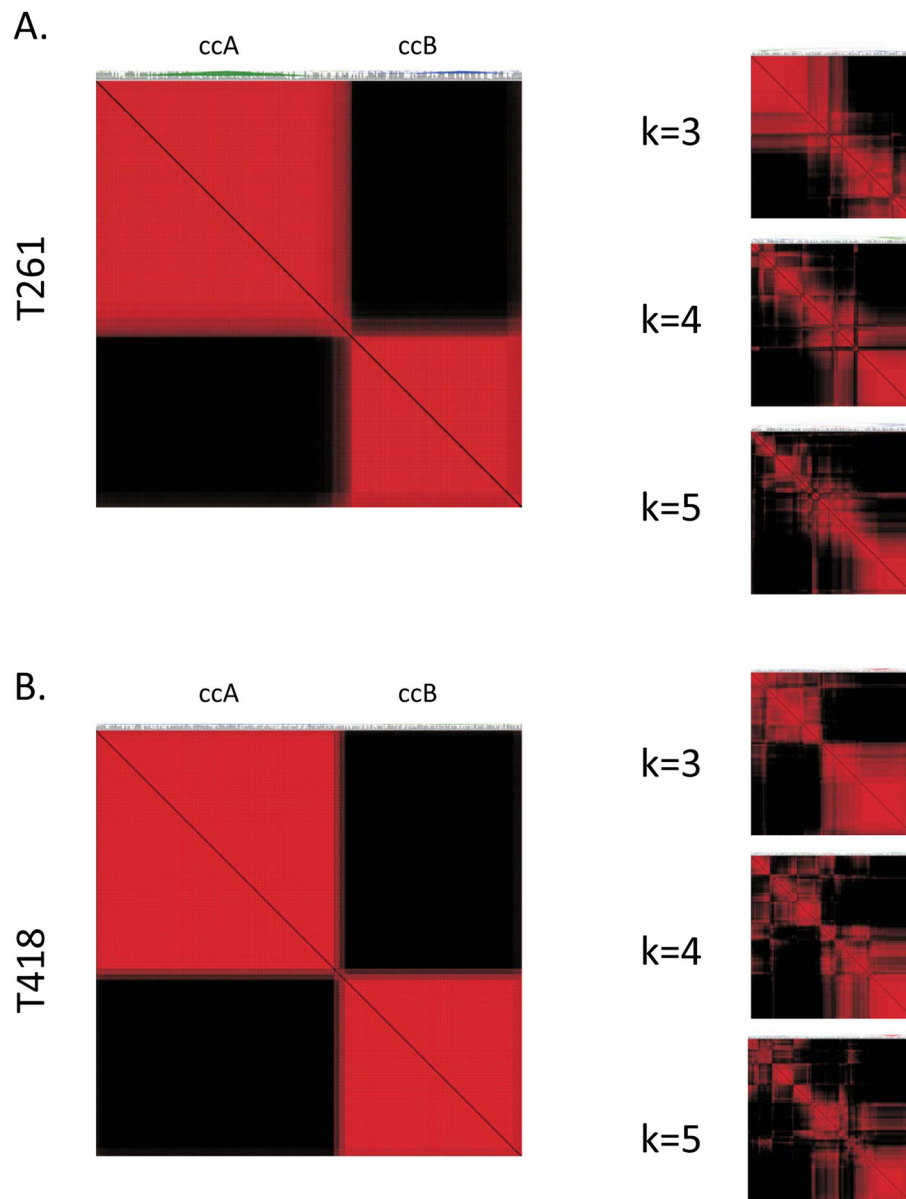
**Fig. 1.** The T480 meta-array is dominated by two clusters, ccA and ccB, but displays a third cluster. Principal component plots showing effective batch effect removal. (a) Before and (b) after batch effect removal in Partek for T480 meta-array. (c) Using ConsensusCluster, two dominant clusters are present in the T480 meta-array, identified as ccA and ccB by gene expression patterns. (d) When ConsensusCluster was set to find  $k = 3$  clusters, a solid third cluster appeared; however, (e, f, and g) consensus matrices for T480 meta-arrays do not show more than three definitive clusters. Additional clustering ( $k = 4, 5,$  and  $6$ ) was directed in ConsensusCluster, but only three clean clusters are apparent. In a–g, red identifies the similarity between samples and display samples clustered together across the bootstrap analysis.



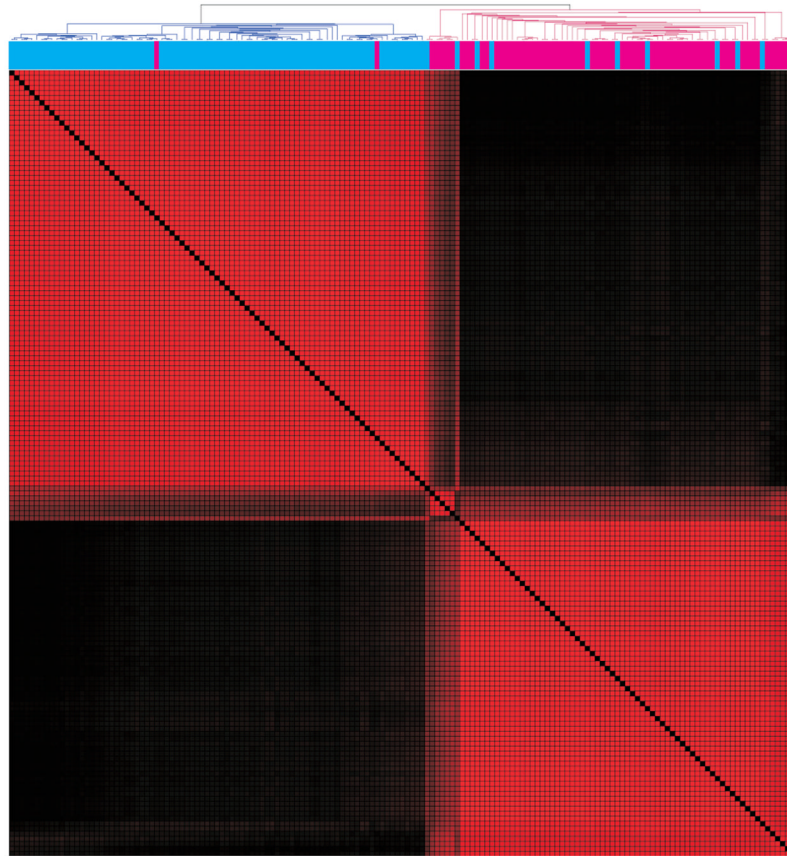
**Fig. 2.** Cluster3 tumors display a distinct metabolic gene profile and are primarily wild type (WT) von Hippel-Lindau (VHL) by gene expression. (a) Single-sample Gene Set Enrichment Analysis (GSEA) displays that Cluster3 tumors overexpress gene-related oxidative phosphorylation and the electron transport chain (ETC) as well as the estrogen-related receptor alpha. (b) Using Biometric Research Branch ArrayTools and gene expression of the Gordan et al tumors, the T480 tumors were assigned as *H1H2*, *H2* only, WT VHL, or unclassified. Although Cluster3 tumors are predominantly WT VHL in expression, ccA and ccB tumors are a mix of *H1H2* and *H2*-only tumors. The heat map is the output of a single-sample GSEA for curated gene sets from the Broad Institute and further demonstrates the strongly different expression pattern of Cluster3 tumors. Yellow represents gene set overexpression; blue represents underexpression. WT VHL = wild type von Hippel-Lindau.



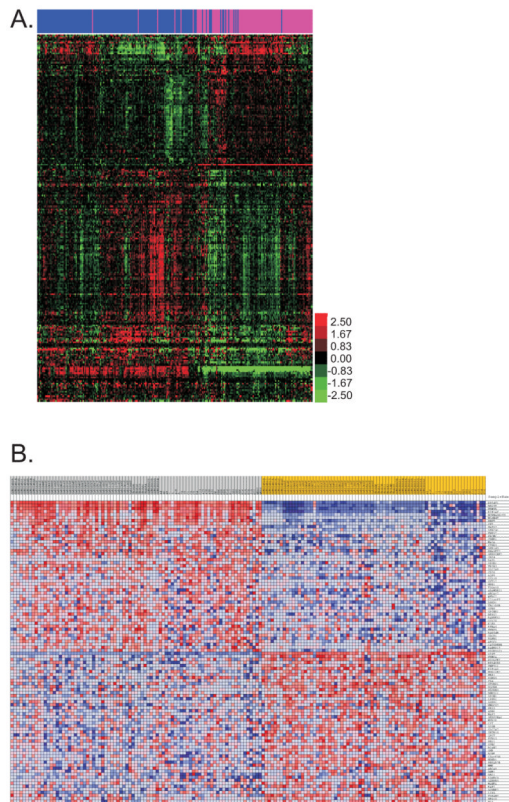
**Fig. 3.** Hematoxylin and eosin staining for ccA, ccB, and Cluster3 tumors. Representative diagnostic images for ccA (a; tumor G7) and ccB (b; tumor J08). Cluster3 tumors shown were classified as clear cell yet display a range of dissimilarity from standard morphology for clear cell (d–f; tumors J04, G8, K13, and 17, respectively). All images taken at  $\times 10$  magnification.



**Fig. 4.** Clear cell-only meta-array display only ccA and ccB clusters. Only two clusters, ccA and ccB, remain when (a) T261 and (b) T418 meta-arrays undergo consensus clustering, even when additional clusters are imposed.



**Fig. 5.** ccA tumors naturally subdivide by gender. When ccA tumors in the meta-array are clustered in an unsupervised manner using ConsensusCluster, two clusters appear most strongly dominated by a gender signature. The T261 meta-array ccA tumors are shown. Red areas identify the similarity between samples and display samples clustered together across the bootstrap analysis. Males are color-coded in blue, females in pink.



**Fig. 6.** Expression analysis identifies genes differentially expressed in tumors based on gender. (a) Heat map showing genes significantly different by Significance Analysis of Microarrays gene set analysis at false discovery rate  $<0.01$ . Because there is a mild gender bias between the ccA and ccB subtypes, for visualization, genes that were significantly different between the subtypes were removed. Males are color-coded in blue, females in pink. Red indicates gene overexpression. (b) Results from Gene Set Enrichment Analysis (Broad Institute) comparing the two subclusters of ccA identified in Figure 1 show similar genes expressed as in the full meta-array gender analysis. The male-dominated cluster is colored gray, whereas the female-dominated cluster is colored yellow. Red indicates gene set overexpression.

**Table 1**

Gene expression studies included in tumor meta-arrays: studies and tumor numbers included for the compiling of the T480 meta-array\*

Study	GEO accession	Included tumors	Published probes
Beroukhim et al [7]	GSE14994	28	22 215
Brannon et al [5]	GSE16449	51	29 858
Newly prepared	–	44	34 293
Dalgliesh et al [9]	GSE17895	135	17 726
Gordan et al [6]	GSE11904	21	13 067
Wuttig et al [8]	GSE22541	24	20 372
Zhao et al [4]	GSE3538	177	19 957
Total	–	480	6386

GEO = Gene Expression Omnibus.

\* The T418 meta-array included the same studies, only excluding tumors not clearly designated as clear cell; the T261 did not include Zhao et al to increase genomic coverage.



Table 2

## Genes differentially expressed by gender\*

<b>M &gt; F</b>	
Cell adhesion	<i>CD2, CD72, CDHRI, CORO1A, CSF3R, CUZD1, KALI, LPXN, NLGN4Y, TGFBI</i>
Cytoskeleton	<i>BASPI, CORO1A, CSTA, HAUS7, MYO1F, MYO9B, NEFL, SYNPO, TREX2, TYK2</i>
GTPase regulation	<i>ARHGAP4, CYTH4, FMNLI, GMIP, LPAR2, MAP4K1, MYO9B, RIN3, SIPA1</i>
Immune response	<i>ADAM28, ADAMDEC1, APOBEC3C, APOC1, C1QB, C1R, C1RL, C1S, C3, CCL18, CD2, CD70, CDHRI, CFB, CMKLR1, CORO1A, CP, CSF3R, CTSZ, CUZD1, DDX3Y, DDX43, DPEP2, FCGR1B, FCGR3B, HAUS7, HCK, HPSE, IGSF6, IL10RA, IL12RB2, IL1R2, IL3RA, ITPR3, LILRB1, LPAR2, LRRC41, LTB, MAN2B1, MICB, MMP7, MYO1F, NCF1, NEFL, NOD2, PILRA, PLA2G5, RBCK1, RNASSET2, ROBO3, SASH3, SERPINF2, SLAMF8, TCIRG1, TNFSF14, TREM2, TREX2, WFDC2</i>
MAPK regulation	<i>LPAR2, LTB, MAP4K1, NOD2, TRIB3</i>
Metalloprotease activity	<i>ADAM28, ADAMDEC1, DPEP2, MMP7</i>
<b>F &gt; M</b>	
Apoptosis regulation	<i>BDNF, CUL3, EAF2, FASTKD5, GLO1, HBXIP, HSPD1, HSPE1, MCF2, NCKAP1, PEG10, PRDX3, SGK1, SGMS1, SGPP1</i>
Catabolism	<i>AKR7A2, ARSD, FBP1, GLUD1, GLUD2, HAO2, HIBCH, IDI1, RDH11</i>
Cell adhesion	<i>CD164, EZR, PVRL3, STXBPI, SGCE</i>
Cell cycle	<i>CIT, CUL3, RNF103, SMC1A, TUBB2A</i>
DNA repair	<i>PTPLAD1, POLE3, PTPLAD1, SMC1A, SUMO1, UBE2N</i>
Mitochondria	<i>ABCD3, ACAA2, ACAT1, AKR7A2, ATP6V1A, ATP7B, C1QBP, C2ORF47, CYP4A11, DDX3X, DNAJB9, EAF2, ETFA, GLUD1, GLUD2, GRIA4, HIBCH, HMGCS2, HSDL2, HSPD1, HSPE1, KDM6A, LRRC59, MED21, MRPS28, OXCT1, PDK4, PEX2, PGRMC1, PRDX3, RNF103, SDHC, SGMS1, SRP9, STXBPI, SUMO1, TMEM38B, TOM1L1, TSPYLI, UQCRH, ZBTB10, ZFX</i>
Oxidation reduction	<i>ADH6, AKR7A2, CYP4A11, ETFA, GLUD1, GLUD2, HAO2, HSDL2, KDM6A, NOX4, PRDX3, RDH11, SDHC, STEAP1, UQCRH</i>
Phosphorylation	<i>ATP6V1A, CIT, FASTKD5, GMFB, HSPD1, HSPE1, PDK4, PRDX3, PRKARIA, PTPLAD1, PTPN13, RCAN1, SGK1, SGMS1, UBE2N, UQCRH</i>
Ubiquitination	<i>CLCN5, CUL3, DPP4, FBXO3, KLHL9, MGAM, PEG10, RNF103, RNF11, SGK1, SUMO1, TOM1L1, UBE2N, WWP1</i>

M = male; F = female; MAPK = mitogen-activated protein kinase.

\* Certain genes were overexpressed in one gender relative to the other in clear cell renal cell carcinoma and may be important to the biological differences observed in males and females. Genes significant (false discover rate <0.01) in both T261 and T418 or significant in T261 but not present in the T418 meta-array were categorically labeled using Database for Annotation, Visualization, and Integrated Discovery. Analyses were performed on a meta-array of normal samples, and none of the above genes was significant in that analysis.

**Table 3**

Catabolic versus immune-related gene ontologies with gender expression bias\*

Gender bias	Gene ontology	T261	T418
<b>M &gt; F</b>			
	ADAPTIVE_IMMUNE_RESPONSE	X	X
	ADAPTIVE_IMMUNE_RESPONSE_GO_0002460	X	X
	CELL_ACTIVATION	X	X
	CELLULAR_DEFENSE_RESPONSE	X	X
	CHEMOKINE_ACTIVITY	X	X
	CHEMOKINE_RECEPTOR_BINDING	X	-
	CYTOKINE_ACTIVITY	-	X
	CYTOKINE_AND_CHEMOKINE_MEDIATED_SIGNALING_PATHWAY	X	-
	CYTOKINE_BINDING	X	X
	CYTOKINE_BIOSYNTHETIC_PROCESS	X	X
	CYTOKINE_METABOLIC_PROCESS	X	X
	CYTOKINE_PRODUCTION	X	X
	CYTOKINE_SECRETION	X	-
	DEFENSE_RESPONSE	X	X
	DEFENSE_RESPONSE_TO_BACTERIUM	X	X
	HEMATOPOIETIN_INTERFERON_CLASSD200_DOMAIN_CYTOKINE_RECEPTOR_ACTIVITY	X	-
	HUMORAL_IMMUNE_RESPONSE	X	-
	IMMUNE_RESPONSE	X	X
	IMMUNE_SYSTEM_DEVELOPMENT	X	X
	IMMUNE_SYSTEM_PROCESS	X	X
	INFLAMMATORY_RESPONSE	X	X
	INNATE_IMMUNE_RESPONSE	X	-
	INTERLEUKIN_BINDING	X	-
	INTERLEUKIN_RECEPTOR_ACTIVITY	X	-
	LEUKOCYTE_ACTIVATION	X	X
	LEUKOCYTE_DIFFERENTIATION	X	-
	LYMPHOCYTE_ACTIVATION	X	X
	MYELOID_LEUKOCYTE_DIFFERENTIATION	X	-
	POSITIVE_REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS	-	X
	POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS	-	X
	POSITIVE_REGULATION_OF_JNK_ACTIVITY	X	-
	POSITIVE_REGULATION_OF_LYMPHOCYTE_ACTIVATION	X	X
	POSITIVE_REGULATION_OF_T_CELL_ACTIVATION	-	X
	REGULATION_OF_CELL_PROLIFERATION	X	-
	REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS	-	X
	REGULATION_OF_DEFENSE_RESPONSE	X	-
	REGULATION_OF_IMMUNE_EFFECTOR_PROCESS	X	-
	REGULATION_OF_IMMUNE_SYSTEM_PROCESS	X	X

Gender bias	Gene ontology	T261	T418
	REGULATION_OF_LYMPHOCYTE_ACTIVATION	X	X
	REGULATION_OF_T_CELL_ACTIVATION	-	X
	REGULATION_OF_T_CELL_PROLIFERATION	-	X
	RESPONSE_TO_BACTERIUM	X	-
	RESPONSE_TO_EXTERNAL_STIMULUS	X	-
	RESPONSE_TO_OTHER_ORGANISM	X	-
	RESPONSE_TO_VIRUS	X	-
	T_CELL_ACTIVATION	X	X
	T_CELL_PROLIFERATION	-	X
	RNA_CATABOLIC_PROCESS	X	-
<b>F &gt; M</b>			
	AEROBIC_RESPIRATION	X	-
	BIOPOLYMER_CATABOLIC_PROCESS	X	-
	CATABOLIC_PROCESS	X	-
	CELLULAR_CATABOLIC_PROCESS	X	-
	CELLULAR_MACROMOLECULE_CATABOLIC_PROCESS	X	-
	CELLULAR_PROTEIN_CATABOLIC_PROCESS	X	X
	CELLULAR_RESPIRATION	X	-
	ENERGY_DERIVATION_BY_OXIDATION_OF_ORGANIC_COMPOUNDS	X	-
	MACROMOLECULE_CATABOLIC_PROCESS	X	-
	MITOCHONDRIAL_ENVELOPE	X	-
	MITOCHONDRIAL_INNER_MEMBRANE	X	-
	MITOCHONDRIAL_MEMBRANE	X	-
	MITOCHONDRIAL_MEMBRANE_PART	X	-
	MITOCHONDRIAL_OUTER_MEMBRANE	X	-
	MITOCHONDRIAL_PART	X	-
	MITOCHONDRIAL_RESPIRATORY_CHAIN	X	-
	MITOCHONDRIAL_TRANSPORT	X	-
	MITOCHONDRION	X	-
	OXIDOREDUCTASE_ACTIVITY	X	-
	OXIDOREDUCTASE_ACTIVITY_ACTING_ON_THE_ALDEHYDE_OR_OXO_GROUP_OF_DONORS	X	X
	OXIDOREDUCTASE_ACTIVITY_ACTING_ON_THE_CH_CH_GROUP_OF_DONORS	X	-
	PROTEIN_CATABOLIC_PROCESS	X	-

M = male; F = female; X = significant; - = not significant.

\* Catabolic gene ontologies were disproportionately associated with females, whereas immune-related gene ontologies were disproportionately associated with males ( $p < 0.05$  for all significant gene ontologies). Analyses were performed on normal meta-array, and none of the above gene ontologies demonstrated the same gender bias.