

**SUPPLEMENTAL DATA: Figures and Tables****Figure 4 (supplemental). Differentially expressed genes were validated by QPCR**

The gene expression of *IGFBP1*, *IGFBP 3*, *CTGF*, *AKT*, *FRAP*, *MYC*, *NF-kB*, *HK1*, *SIRT7*, *PHD1*, was validated by QPCR. The gene expression of *PHD2* and *PHD3* was quantified as well.

**Table 4 (supplemental). The RRR 1,325 genes expression data and specific functional gene-clusters**

1,325 unique genes were identified in the current microarray dataset. The gene expression is presented as up or down from normal-ischemic kidneys. Two separate groups of microarray experiments were conducted, and the results were subsequently normalized to eliminate systematic bias. The first group consisted of normal and ischemic tissues, as well as and 1 and 2 days post-injury. The second group consisted of normal kidneys and 5 and 14 days post-injury. The data from days 1 and 2 were normalized by the mean of the normal-ischemic group, and the data from days 5 and 14 by the mean of the corresponding normal kidney. The genes were further clustered according to RCC vs. normal kidney; renal cell culture hypoxia responsive genes vs. normoxia; HIF regulated genes; VHL, IGF1, MYC, NF-kB pathway genes; purine pathway genes; gene expression following renal ischemia reperfusion and/or acute renal failure (ARF) vs. normal tissue; and gene expression in response to serum (1, 2).

**Table 5 (supplemental). An ontology analysis in timely dependent fashion: distinct and common ontologies**

**A.** The differentially expressed genes were clustered according to their pattern of expression as early, late or continually RRR. Functional ontology analysis was performed ( $p < 0.05$ ). The presented ontologies are the ontology core and are hyperlinked to EMBL-EBI. The average RRR expression ( $\log_2$ ) of each ontology is presented in a green to red scale; green down-regulated, red up-regulated. The numbers and average RRR expression of up- and down- regulated genes, the category p-value and enrichment are shown as well.

**B.** The genes in the three phases of renal regeneration and the concordant and discordant genes are analyzed for GO (summary sheets). These genes were crossed with the data from supplemental Table 4 (cross sheets); green down-regulated and red up-regulated in RRR.

**Table 6 (supplemental). The differentially expressed genes in both RRR and RCC exhibited distinct ontologies for the concordance vs. discordance genes**

The differentially expressed genes in both RRR and RCC were clustered according to their concordance vs. discordant change. Functional ontology was analysis performed ( $p < 0.05$ ). The ontologies are hyperlinked to EMBL-EBI. The average RRR expression of each ontology is presented in a green to red scale; green down-regulated, red up-regulated. The number and average RRR expression of genes up- / down- regulated in both RRR and RCC, the category p-value and enrichment are also given (the expression direction and values are as in RRR, relative to the normal kidney).

**Table 7 (supplemental). The significance of gene in the various expression groups: patterns, trends and pathways**

The significance of gene in the various expression patterns of early, late, continues, pathways and the concordant or discordant groups was analyzed by using the chi square test. See methods for further explanation.

**Table 8 (supplemental). The RRR genes in non-probabilistic GO ontologies**

The comprehensive probabilistic analysis may fail to capture many key aspects of the concordant and discordant gene functions. Therefore, we also categorized the genes into gene-by-gene, non-probabilistic GO.

**Table 9 (supplemental). An ontology analysis of the concordant and discordant genes in pathway dependent fashion: distinct and common ontologies**

The concordantly and discordantly differentially expressed genes were clustered according to their regulation by the pathways of VHL, hypoxia, HIF, IGF1, MYC, p53 and NF-kB. Functional ontology was analysis performed ( $p < 0.05$ ).

**SUPPLEMENTAL DATA: Text**

**Supplemental discussion on discordant genes- biological processes that differentiate RRR from RCC**

*Fatty acid metabolism*

Fatty acid metabolism plays a major role in cancer. Our study found that two fatty acid metabolic enzymes, Acyl-Coenzyme A oxidase 1 (*ACOX1/1.3.3.6*)

and Carnitine PalmitoylTransferase 1A (liver) (*CPT1A/ 2.3.1.21*) are up-regulated in RCC, but down-regulated during the late pattern or continually during RRR. The over-expression of both enzymes may increase the levels of intracellular H<sub>2</sub>O<sub>2</sub> which may drive the carcinogenic process or influence signal transduction pathways (3).

The organic cation transporter, solute carrier family 22 (*SLC22A1*), is critical for the elimination of many endogenous small organic cations, as well as a wide range of drugs and environmental toxins, in kidney and other tissues. *SLC22A1* is up-regulated in RCC, but down-regulated in RRR (supplemental Table 4). It may play a role in eliminating toxins from carcinoma cells due to increased metabolism (4).

#### *mRNA maturation, transcription and post-translational*

poly(A) polymerase (*PAPOLA/PAP*) is down-regulated throughout RRR and up-regulated during RCC (supplemental Table 4). This gene is of particular interest because increased PAPOLA activity is associated with rapidly proliferating cells. It is a strong anti-apoptotic protein and important marker for poor prognosis in leukemia and breast cancer patients (5, 6). The guanine-rich sequence factor 1 (*GRSF1*) is down-regulated late during RRR and up-regulated during RCC, possibly by hypoxia (7). GRSF1 binds cellular and viral mRNAs via a guanine-rich sequence motif in the 5' UTR, thereby recruiting these mRNAs to polyribosomes (8). GRSF1 could possibly regulate the expression of guanine-rich transcripts, as described for the EVI1 proto-oncogene (9, 10).

The nuclear orphan receptor-binding retinoic acid response 6 (*NR2F6/EAR2*) is up-regulated late during RRR and down-regulated during RCC (supplemental Table 4), (11). *NR2F6* is a negative modulator of renin transcription, possibly explaining high levels of renin and prorenin associated with RCC (12). The nuclear receptor coactivator4 (*NCOA4/ ARA70*) is down-regulated late during RRR and up-regulated in RCC (supplemental Table 4);

(13). NCOA4 is a ligand-dependent androgen receptor (AR)-associated protein that enhances AR transcriptional activity 10-fold in prostate cancer cells in the presence of dihydrotestosterone or testosterone, but not hydroxyflutamide. NCOA4 is AR-specific and does not stimulate the estrogen receptor in human prostate cancer cells (14). Collectively, these data suggest that transcriptional and post-translational control may be deregulated tumor cells.

#### *DNA repair*

Several DNA repair and/or stress response genes are discordant in RRR and RCC. These genes include *SMC1L1*, *TOP3B* and *FRAP1* and possibly *SIRT7*. *SMC1L1* (structural maintenance of chromosomes 1-like 1) is up-regulated early during RRR and down-regulated during RCC (Table 3 and supplemental Table 4). *SMC1L1* is essential for sister chromatid cohesion in yeast cells undergoing mitosis and may play a role in DNA repair (15). *TOP3B* (topoisomerase III beta) is down-regulated early during RRR and up-regulated during RCC (Table 3 and supplemental Table 4). *TOP3B* interacts with DNA helicase *SGS1* and plays a role in DNA recombination, cellular aging, and the maintenance of genome stability (16).

#### *Heparin binding*

Five discordant genes (*CTGF*, *THBS1*, *VEGF*, *APOE* and *LPL*) are interacting selectively with heparin. The first 3 genes are involved in angiogenesis (Tables 2 and supplemental Table 5b). Heparin-binding angiogenic proteins are stored as a complex with heparan sulfate in the microenvironment of tumors. These proteins are released and can induce new capillary growth when heparan sulfate is degraded by heparanase (17).

#### *Autosomal dominant polycystic kidney disease (ADPKD) and RCC*

Patients suffering from chronic renal failure tend to develop cystic kidney disease which can progress to RCC (18). Polycystic kidney disease 1 (*PKD1*) is down-regulated early during RRR and up-regulated during RCC (supplemental

Table 4); (7). *PKD1* encodes a glycoprotein that may function as an integral membrane protein involved in cell-cell/matrix interactions, and may modulate intracellular calcium homeostasis and other signal-transduction pathways. *PKD1* plays a role in renal tubular development, and mutations in *PKD1* account for 85% of autosomal dominant polycystic kidney disease (ADPKD), (19, 20). It is possible that up-regulation of *PKD1* in RCC is consequence of the pathologic tumor hypoxia (7). Alternatively, *PKD1* may play some unknown function in RCC.

#### *The HIF-VHL pathway (continued)*

*UBE2V1/CIR1*, a variant E2 ubiquitin-conjugating enzyme, is down-regulated during RRR and up-regulated in immortalized renal cells and in human tumor cell lines (21). Further studies would be needed to define the relationship between *UBE2V1/CIR1* and HIF1a or the E2 ubiquitin-conjugating enzymes CUL2, CUL5, CUL1 and CDC34. Another discordantly expressed protein ligase is zinc finger protein 144 (*ZNF144/PCGF2/MEL18*). *ZNF144* is up-regulated in late RRR and discordantly down-regulated in RCC (supplemental Tables 4 and 8); (13). The protein encoded by *ZNF144* contains a RING finger motif and is similar to the polycomb group (PcG) gene products. The human PcG protein, Pc2, is a SUMO E3 (small ubiquitin-related modifier E3); (22). *ZNF144* is proposed as a negative regulator of self-renewal of the hematopoietic stem cells and promoter of their differentiation. It also appears to be a tumor suppressor (23). Further studies focused on the role of those ligases in RCC and the HIF pathway are warranted (24-26).

#### *The IGF1 pathway (continued)*

*IGFBP8/CTGF*, a member of the CCN family, is also differentially up-regulated late during RRR and down-regulated in RCC. *CTGF* binds *IGF1*, albeit with relatively low affinity compared with classical *IGFBPs*. *CTGF* and *IGF1* cooperate in up-regulating type I and III collagens in human renal fibroblasts. The synergy between *CTGF* and *IGF1* might be involved in glucose-induced matrix accumulation, because both factors are induced by hyperglycemia (27). *CTGF* is

a major factor in fibrotic disease, is down stream of TGF $\alpha$  and may inhibit metastasis and invasion (28-30).

### **Methodological considerations related to data integration, quality control, data cataloging and data validation**

#### *Implementation of comparative biology in the current study*

RRR in human though common (*eg.*, kidney transplantation); (31) is not amenable for obtaining samples at different times during RRR. In recent years, mouse (and other) model systems have shed new light on the nature and treatment of human RRR. Included have been physiological, pharmacological, global gene expression and gene inactivation studies (32, 33) Therefore, the changes in RRR gene expression were measured in mouse model. To the best of our knowledge there is no mouse model for sporadic RCC. Therefore, to compare RRR and RCC we performed a careful comparative biology analysis of differential gene expression in RRR and RCC. This required integrated data from different organisms, tissue pathologies, methods and authors (34). The interspecies comparison of gene expression of mouse RRR with human RCC was feasible by using the normal tissue in each original publication as a reference point and thus the comparison was indirect (*i.e.* not RRR vs. RCC). The significance of the differentially expressed genes was as offered by the authors. The feasibility of the comparison was supported by the findings that both the RCC and the RRR process are predominantly found in the proximal tubules which make the bulk of the kidney tissue (Figures 1), (35-39).

There are two basic approaches to address the complexity of cancer. One is to reduce complexity through analysis of experimental models (cell lines or animal models) to characterize fundamental processes or function of single genes. Another is to integrate large data sets, to yield a model for tumor development and behavior (34). Each approach has its own advantages and disadvantages (40-42). The second approach, involving integration of large data

sets, is challenging in part because only a limited number of samples, such as tumors, preneoplastic tissues or wound healing samples, can be analyzed in a given study. This makes data interpretation and model development difficult, given the large amount of heterogeneity between human tumors (34). Thus the analytical approach we took was to correlate mouse experimental RRR model data with the very extensive data reported on RCC. Comparisons across biological systems are commonly done for example using tissue culture serum response to predict cancer survival (1, 2). Further, the power of this comparative-analysis approach is also exemplified by a comparison of the RRR literature with the current experimental RRR dataset. Of the 91 genes appeared on both lists, 89% were differentially expressed in full agreement (up or down), despite the difference in organisms (human, rat, mouse) and methods (supplemental Table 4). Therefore, qualitative comparative-analysis and data integration is plausible if the normal tissue is used as a reference point.

To reduce the noise in the datasets, the differential RCC gene expression was catalogued only qualitatively (not quantitatively), as expressed up or down from normal tissue (supplemental table 4). The comparison to RRR gene expression was qualitative comparison as well. The comparison resulted in three subsets, (1) genes that were differentially expressed in RRR as in RCC; (2) genes that were discordantly expressed in RRR versus RCC; (3) genes with conflicting data as per to their expression in RCC. The last group was not analyzed further. All these subsets include noise, which is a result of differences in tissue pathologies, methods and authors and is assumed to be distributed homogenously in both sets of the concordant and discordant genes.

This approach was justified by having large data samples and simply assuming that the distribution we have sampled is approximately normal. This enabled us to take into consideration and integrate the extensive heterogeneous plethora of information on RCC, gathered through methods as microarray, northern and qPCR.

### *Validation of the microarray dataset*

A global knowledge step toward constructing a RRR systems biology network model is to build a comprehensive RRR expression database. Therefore we reviewed the evidence reported in the literature on differentially expressed genes in RRR and the relevant pathways and cross-compared them with the current study (supplemental Table 4). Of the 1,325 RRR differentially expressed genes in the current study, the expression of 91 genes was previously compared with normal kidney. The qualitative expression of 89% of the 91 genes was in full agreement and only 11% was in qualitative conflict that included the genes: NID, NRP1, ZFP36L1, TNC, MAPK1, HSPD1, HK1, NEDD4, CASP1 and UK114. These results were despite the difference in organisms (human, mouse) and methods (supplemental Table 4). We further tested the validity of the data by RT-QPCR of IGFBP1, IGFBP 3, CTGF, AKT, FRAP, MYC, NF- $\kappa$ B, HK1, SIRT7 and PHD2 (EGLN1). The expression of *PHD2* was down-regulated in early and late regenerating kidney in comparison to resting/normal kidney. Similar expression patterns were repeated with two other related prolyl hydroxylases, PHD1 and PHD3 that were down-regulated as well (supplemental Figure 4). Lastly, The MiB-1 high expression at 2 days was in full agreement with the array results (supplemental Table 4).

### **SUPPLEMENTAL DATA: References**

1. Chang HY, Sneddon JB, Alizadeh AA, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004;2:E7.
2. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* 2005;102:3738-43.
3. Okamoto M, Reddy JK, Oyasu R. Tumorigenic conversion of a non-tumorigenic rat urothelial cell line by overexpression of H<sub>2</sub>O<sub>2</sub>-generating peroxisomal fatty acyl-CoA oxidase. *Int J Cancer* 1997;70:716-21.

4. Shu Y, Leabman MK, Feng B, et al. Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1. *Proc Natl Acad Sci U S A* 2003;100:5902-7.
5. Balatsos NA, Havredaki M, Tsiapalis CM. Early 5-fluorouracil-induced changes of poly(A) polymerase in HeLa and WISH cells. *Int J Biol Markers* 2000;15:294-9.
6. Stetler DA, Rose KM, Jacob ST. Anti-poly(A) polymerase antibodies in sera of tumor-bearing rats and human cancer patients. *Proc Natl Acad Sci U S A* 1981;78:7732-6.
7. Jiang Y, Zhang W, Kondo K, et al. Gene expression profiling in a renal cell carcinoma cell line: dissecting VHL and hypoxia-dependent pathways. *Mol Cancer Res* 2003;1:453-62.
8. Kash JC, Cunningham DM, Smit MW, et al. Selective translation of eukaryotic mRNAs: functional molecular analysis of GRSF-1, a positive regulator of influenza virus protein synthesis. *J Virol* 2002;76:10417-26.
9. Bartholomew C, Clark AM. Induction of two alternatively spliced evi-1 proto-oncogene transcripts by cAMP in kidney cells. *Oncogene* 1994;9:939-42.
10. Takahashi S, Licht JD. The human promyelocytic leukemia zinc finger gene is regulated by the Evi-1 oncoprotein and a novel guanine-rich site binding protein. *Leukemia* 2002;16:1755-62.
11. Gieseg MA, Cody T, Man MZ, Madore SJ, Rubin MA, Kaldjian EP. Expression profiling of human renal carcinomas with functional taxonomic analysis. *BMC Bioinformatics* 2002;3:26.
12. Liu X, Huang X, Sigmund CD. Identification of a nuclear orphan receptor (Ear2) as a negative regulator of renin gene transcription. *Circ Res* 2003;92:1033-40.
13. Boer JM, Huber WK, Sultmann H, et al. Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. *Genome Res* 2001;11:1861-70.
14. Yeh S, Chang C. Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. *Proc Natl Acad Sci U S A* 1996;93:5517-21.
15. Sumara I, Vorlaufer E, Gieffers C, Peters BH, Peters JM. Characterization of vertebrate cohesin complexes and their regulation in prophase. *J Cell Biol* 2000;151:749-62.
16. Li W, Wang JC. Mammalian DNA topoisomerase IIIalpha is essential in early embryogenesis. *Proc Natl Acad Sci U S A* 1998;95:1010-3.
17. Zcharia E, Metzger S, Chajek-Shaul T, et al. Molecular properties and involvement of heparanase in cancer progression and mammary gland morphogenesis. *J Mammary Gland Biol Neoplasia* 2001;6:311-22.
18. Li MK, Choy DK, Yip SK. Renal cell carcinoma in patients with chronic renal failure. *Ann Acad Med Singapore* 1999;28:512-5.
19. Kleymenova E, Ibraghimov-Beskrovnaya O, Kugoh H, et al. Tuberin-dependent membrane localization of polycystin-1: a functional link between polycystic kidney disease and the TSC2 tumor suppressor gene. *Mol Cell* 2001;7:823-32.

20. Hughes J, Ward CJ, Peral B, et al. The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat Genet* 1995;10:151-60.
21. Ma L, Broomfield S, Lavery C, Lin SL, Xiao W, Bacchetti S. Up-regulation of CIR1/CROC1 expression upon cell immortalization and in tumor-derived human cell lines. *Oncogene* 1998;17:1321-6.
22. Kagey MH, Melhuish TA, Wotton D. The polycomb protein Pc2 is a SUMO E3. *Cell* 2003;113:127-37.
23. Kanno M, Hasegawa M, Ishida A, Isono K, Taniguchi M. mel-18, a Polycomb group-related mammalian gene, encodes a transcriptional negative regulator with tumor suppressive activity. *Embo J* 1995;14:5672-8.
24. Tojo M, Matsuzaki K, Minami T, et al. The aryl hydrocarbon receptor nuclear transporter is modulated by the SUMO-1 conjugation system. *J Biol Chem* 2002;277:46576-85.
25. Ayaydin F, Dasso M. Distinct in vivo dynamics of vertebrate SUMO paralogues. *Mol Biol Cell* 2004;15:5208-18.
26. Shao R, Zhang FP, Tian F, et al. Increase of SUMO-1 expression in response to hypoxia: direct interaction with HIF-1alpha in adult mouse brain and heart in vivo. *FEBS Lett* 2004;569:293-300.
27. Lam S, van der Geest RN, Verhagen NA, et al. Connective tissue growth factor and igf-I are produced by human renal fibroblasts and cooperate in the induction of collagen production by high glucose. *Diabetes* 2003;52:2975-83.
28. Leask A, Abraham DJ. TGF-beta signaling and the fibrotic response. *Faseb J* 2004;18:816-27.
29. Kim KH, Park GT, Lim YB, et al. Expression of connective tissue growth factor, a biomarker in senescence of human diploid fibroblasts, is up-regulated by a transforming growth factor-beta-mediated signaling pathway. *Biochem Biophys Res Commun* 2004;318:819-25.
30. Chang CC, Shih JY, Jeng YM, et al. Connective tissue growth factor and its role in lung adenocarcinoma invasion and metastasis. *J Natl Cancer Inst* 2004;96:364-75.
31. Star RA. Treatment of acute renal failure. *Kidney Int* 1998;54:1817-31.
32. Bouchard M. Transcriptional control of kidney development. *Differentiation* 2004;72:295-306.
33. Heyman SN, Lieberthal W, Rogiers P, Bonventre JV. Animal models of acute tubular necrosis. *Curr Opin Crit Care* 2002;8:526-34.
34. Hanash S. Integrated global profiling of cancer. *Nat Rev Cancer* 2004;4:638-44.
35. Price PM, Megyesi J, Safirstein RL. Cell cycle regulation: repair and regeneration in acute renal failure. *Semin Nephrol* 2003;23:449-59.
36. Linehan WM, Zbar B. Focus on kidney cancer. *Cancer Cell* 2004;6:223-8.
37. Cheung C, Vesey D, Cotterill A, et al. Altered messenger RNA and protein expressions for insulin-like growth factor family members in clear cell and papillary renal cell carcinomas. *Int J Urol* 2005;12:17-28.

38. Avery AK, Beckstead J, Renshaw AA, Corless CL. Use of antibodies to RCC and CD10 in the differential diagnosis of renal neoplasms. *Am J Surg Pathol* 2000;24:203-10.
39. McGregor DK, Khurana KK, Cao C, et al. Diagnosing primary and metastatic renal cell carcinoma: the use of the monoclonal antibody 'Renal Cell Carcinoma Marker'. *Am J Surg Pathol* 2001;25:1485-92.
40. Tapon N. Modeling transformation and metastasis in *Drosophila*. *Cancer Cell* 2003;4:333-5.
41. Adryan B, Decker HJ, Papas TS, Hsu T. Tracheal development and the von Hippel-Lindau tumor suppressor homolog in *Drosophila*. *Oncogene* 2000;19:2803-11.
42. Rangarajan A, Weinberg RA. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat Rev Cancer* 2003;3:952-9.