*Drosophila yakuba* – *Tsc1*

Bailey Lose1Ω, Abigail Myers1Ω, Ian Alberts2, Savanah Fondse3, Joyce Stamm2, James Youngblom3, Chinmay P. Rele1Ω

1. The University of Alabama, Tuscaloosa, AL 35487
2. University of Evansville, Evansville, IN 47722
3. California State University Stanislaus, Turlock, CA 95382

Ω.  Corresponding Author

|  |  |
| --- | --- |
| Species | *Drosophila yakuba* |
| NCBI Taxonomy ID | 7245 |
| NCBI Assembly ID | dyak\_caf1 |
| Assembly Accession Number | GCA\_000005975.1 |
| UCSC Genome Browser Assembly | May 2011 (WUGSC dyak\_caf1/DyakCAF1) |
| Scaffold ID | chr3R |
| NCBI Scaffold Accession Number | CM000160.2 |
| Gene ID in *Drosophila yakuba* | *dyak\_Tsc1* |
| Gnomon ID at model location | XM\_002099254.2 |
| Locus ID | *LOC6538776* |

|  |  |
| --- | --- |
| Informant Species | *Drosophila melanogaster* |
| Informant Annotation Release | FlyBase Release 6.28 |

|  |  |
| --- | --- |
| Official gene symbol of the ortholog | *Tsc1* |
| Accession number of the ortholog | FBgn0026317 |
| Chromosome for ortholog | Chr3R |

## Description

The hamartoma syndrome known as tuberous sclerosis complex (TSC) is caused by mutations in one or both the *Tsc1* and *Tsc2* genes (Dabora, et. al, 2008). These two genes function together in the insulin signaling pathway and also control cell growth which means they also operate as tumor suppressor genes (Gao, 1970). Benign tumors can also be formed in multiple organs due to a mutation in the *Tsc*1 gene (Potter, Huang, Xu, 2001).

## Synteny

**Protein BLAST:**

For each *D. yakuba* NCBI RefSeq gene prediction in Table S1, we obtained the corresponding protein sequence from NCBI and performed a *blastp* search against the collection of *D. melanogaster* protein sequences in the Reference proteins (refseq\_protein) database. Each *blastp* search from the protein sequence derived from the Gnomon data, provided the same gene found in *D. melanogaster* indicating the focal gene, *Tsc1* in *D. yakuba* is an ortholog to the *Tsc1* gene found in *D. melanogaster.* Locus number LOC6538776 corresponds to both *Tsc1* gene and *CENP-ana* gene but there was only a 30.00% identity for *CENP-ana*  which is considerably lower than 97.00% for *Tsc1*  indicating the locus number corresponds to *Tsc1.* LOC6538778 corresponds to the *Root* gene, LOC6538777 corresponds to the *GatB* gene, LOC6538775 corresponds to the *Sec10* gene, and LOC6538774 corresponds to the *Npc2f* gene.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene Prediction Searched** | **Top Hit Accession** | **Gene ID of Top Hit** | **E-value** | **Percent identity** | **Next best gene ID** | **Next best E-value** | **Next best percent identity** |
| *LOC6538778* | NP\_001287497.1 | *Root* | 0.0 | 97.91% | N/A | N/A | N/A |
| *LOC6538777* | NP\_651217.2 | *GatB* | 0.0 | 95.93% | N/A | N/A | N/A |
| *LOC6538776* | NP\_477415.1 | *Tsc1* | 0.0 | 97.00% | *CENP-ana* | 1.2 | 30.00% |
| *LOC6538775* | NP\_651218.1 | *Sec10* | 0.0 | 99.30% | N/A | N/A | N/A |
| *LOC6538774* | NP\_651219.1 | *Npc2f* | 2e-126 | 98.82% | N/A | N/A | N/A |

Table S1: Summary of the *blastp* search results for the genes surrounding the *D. yakuba Tsc1* ortholog *LOC6538776* against the *D. melanogaster* proteins in the refseq\_protein database. The table shows the *D. melanogaster* gene with the best match based on their E-value and percent identity. We have also included the E-value and percent identity for the second-best match (if any) to provide additional evidence for the ortholog assignments

|  |  |  |  |
| --- | --- | --- | --- |
| **NCBI Gene Symbol** | **RefSeq mRNA Accession Number** | **RefSeq Protein Accession Number** | **Putative *D. melanogaster* Ortholog** |
| *LOC6538778* | XM\_002099265.2 | XP\_002099292.2 | *Root* |
| *LOC6538777* | XM\_002099254.2 | XP\_002099291.2 | *GatB* |
| *LOC6538776* | XM\_002099254.2 | XP\_002099290.1 | *Tsc1* |
| *LOC6538775* | XM\_002099253.2 | XP\_002099289.1 | *Sec10* |
| *LOC6538774* | XM\_002099252.2 | XP\_002099288.1 | *Npc2f* |

Table S2: *Gnomon ID/LOCUS ID association:* this figure shows how the Gnomon IDs and LOCUS IDs are associated with the genes surrounding *Tsc1* in *D. yakuba.*

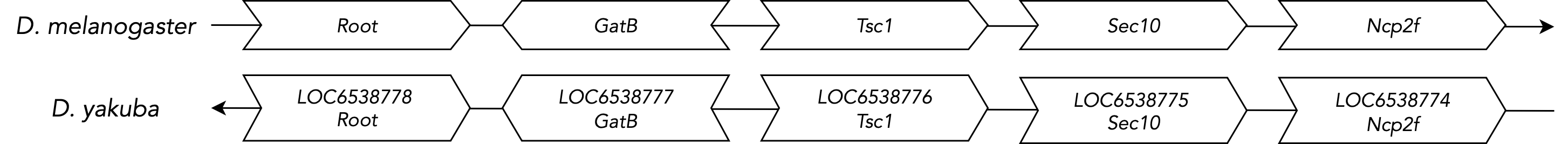


Figure S1: Synteny analysis of the genomic region near *Tsc1* in *D. melanogaster* and *D. yakuba.* The names of genes for *D. melanogaster* correspond to their FlyBase gene symbols. The names of the genes for *D. yakuba* are corresponding to their NCBI gene symbols (top) and the names of the related *D. melanogaster* orthologs identified by *blastp* searches (bottom). For both the *D. yakuba* and *D. melanogaster* tracks, the underlying arrows indicate the strand the focal gene resides on. The direction of the block arrow corresponds to the orientation of the gene relative to *Tsc1*. For the *D. yakuba* track, the color white indicates that the gene is an ortholog of the gene in *D. melanogaster* at the corresponding position with respect to the *Tsc1* gene.

## Model

Untitled picture.png 
Untitled picture.png 


Figure S2: Annotation of *Tsc1* in *D. yakuba* as seen in Apollo. The model labeled DyakCAF1\_Tsc1-PA in the “User-created Annotations” track (highlighted at the top) is shown along with the corresponding *D. melanogaster* protein alignments, NCBI RefSeq Genes, TransDecoder Transcripts, RNA-Seq Coverage for Adult Females, RNA-Seq Coverage for Adult-Males, and Splice Junctions, respectively.

There is one isoform for *Tsc1* in *D. yakuba,* Tsc1-PA, and it contains 6 exons.

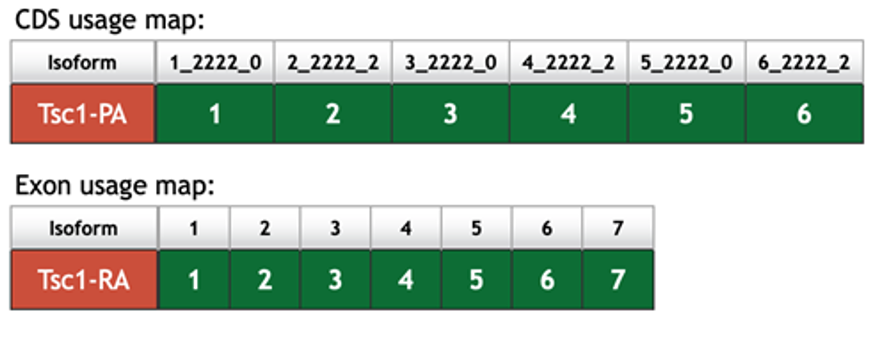


Figure S3: CDS usage map for *dyak\_Tsc1* from *D. melanogaster* in the Gene Record Finder.

## Dot Plot

The diagonal line in the dot plot (Figure S4) indicates the high sequence identity between the *D. melanogaster* Tsc1-PA protein and the *D. yakuba* ortholog. The slight gaps, representing misalignments in protein sequence, along the diagonal line of the dot plot show that there are slight differences between the two protein sequences within the third coding exon, shown within the green box, and the sixth coding exon, shown within the blue box, of Tsc1-PA. There are also two small repeats located within exon 3, represented by the pink box, but these are not significant due to the repeats being present in both sequences.

The alignment has an overall 97.00% identity (Figure S4), due to 33 amino acids differing between the two species, the slope of the line in the dot plot is close to one, which indicates the total length of Tsc1-PA in similar in both *D. melanogaster* and *D. yakuba* indicating an orthologous relationship.

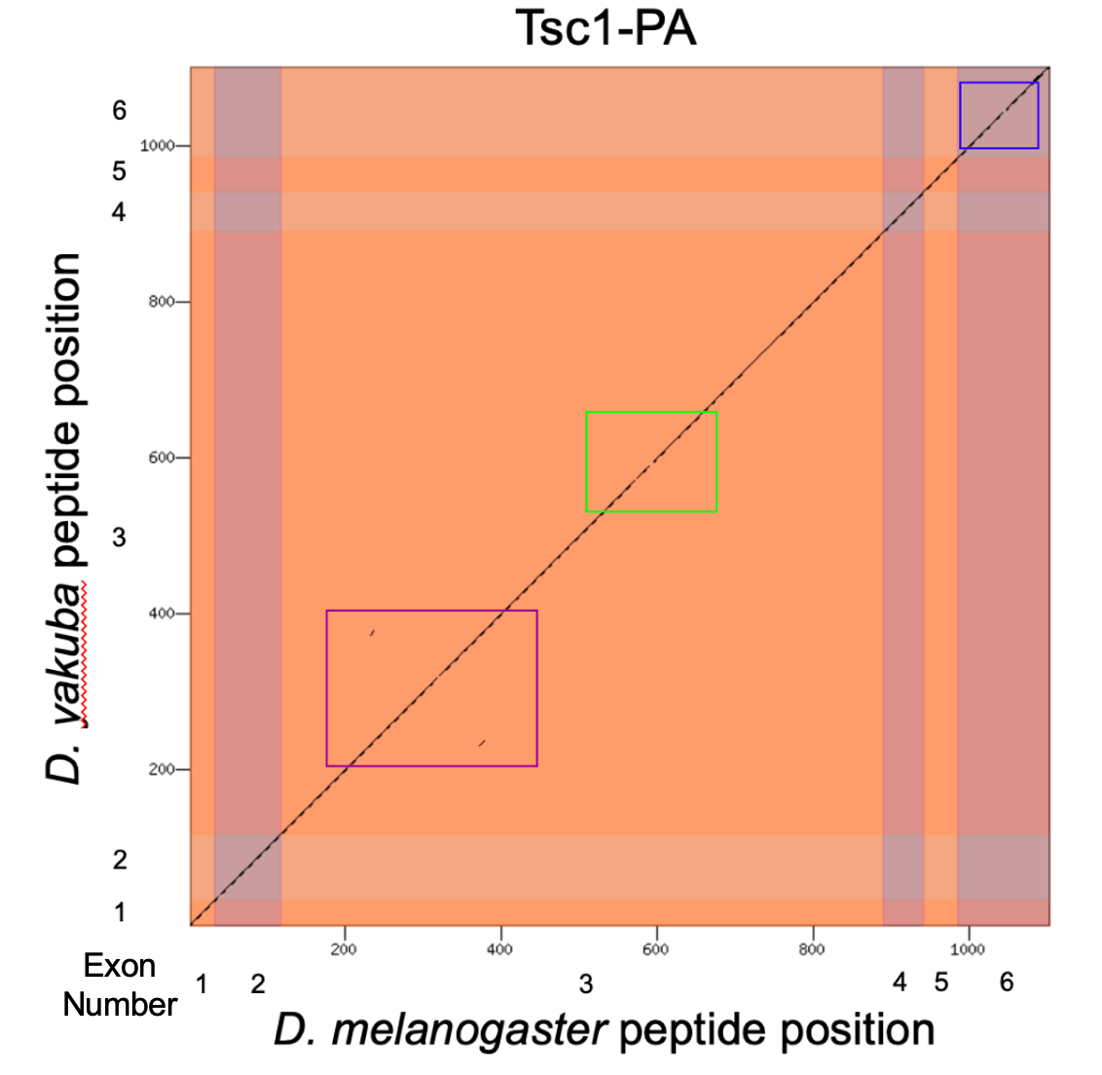


Figure S4: Dot plot comparison of Tsc1-PA in *D. melanogaster* (x-axis) to Tsc1-PA in *D. yakuba* (y-axis).



Figure S5: The protein alignment of the *D. melanogaster* Tsc1-PA protein against the *D. yakuba* ortholog (“Submitted\_Sequence”).

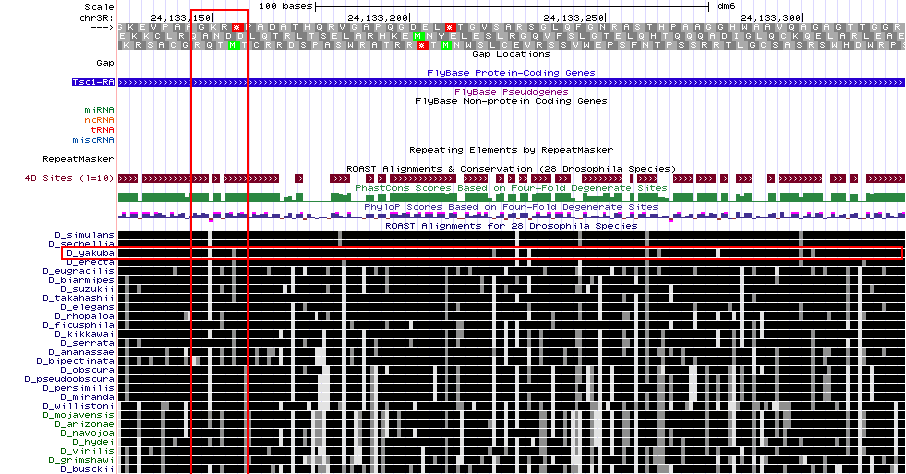


Figure S6: The *ROAST* alignment to the beginning of coding exon 3 of the *D. melanogaster Tsc1* gene contains a misalignment in multiple *Drosophila* species (red brackets). *D. yakuba* is also highlighted in a smaller bracket.

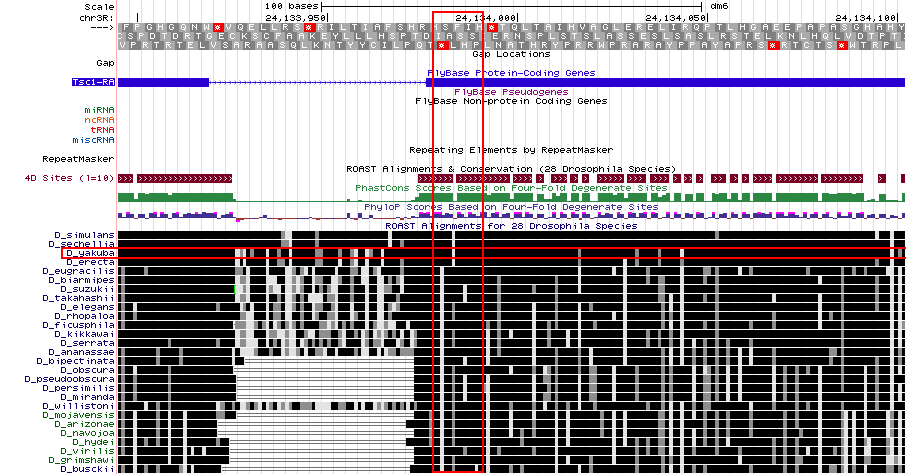


Figure S7: The *ROAST* alignment to the beginning of coding exon 6 of the *D. melanogaster Tsc1* gene contains a misalignment in multiple *Drosophila* species (red brackets). *D. yakuba* is also highlighted in a smaller bracket.

## RNA-Seq Data

The RNA-seq data for the adult male and adult female samples display clean peaks that align with each splice junction and gene prediction (Figure S2). However, there is some RNA-seq coverage for both adult male and adult female before and after the coding sequence for the *Tsc1* gene. This data, which is present in both the NCBI RefSeq Genes and TransDecoder Transcripts tracks, could be due to the 5’ and 3’ UTRs of the mRNA.

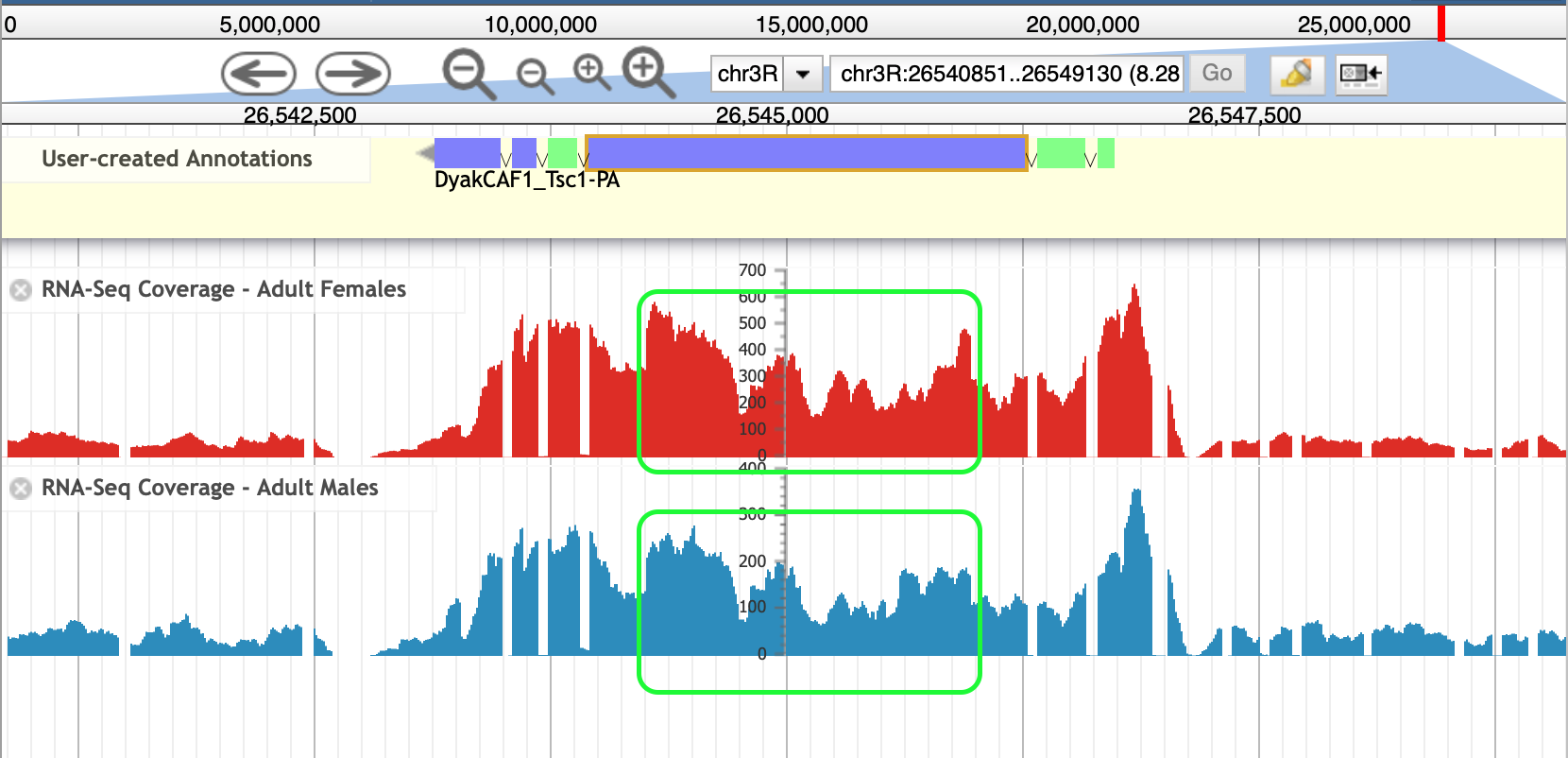


Figure S8: Jump and Drop region presented in green box contained in RNA-Seq Data for Adult Females and Adult Males

There is both a jump up (26545970) and drop down (2654425) in data present in exon 3 for the RNA-Seq Data for both adult males and adult females. The amino acid sequence before the jump up in exon 3 of *D. yakuba* could be considered highly conserved in comparison with *D. melanogaster* due to no differences being present between the amino acids in the two species’ protein alignments. This trend is also true for the RNA-seq coverage after the drop down in data near the end of coding exon 3. Though the majority of the sequence in *D. yakuba* matches in comparison with *D. melanogaster*, the differences show that this region (Figure 8) is less conserved than the flanking regions of exon 3 between *D. yakuba* and *D. melanogaster*.

## Splice Junctions:

|  |  |  |  |
| --- | --- | --- | --- |
| **5’ Exon** | **Splice Junction** | **3’ Exon** | **Splice Junction Score** |
| 1\_2222\_0 | JUNC00081622 | 2\_2222\_2 | 516 |
| 2\_2222\_2 | JUNC00081621 | 3\_2222\_0 | 405 |
| 3\_2222\_0 | JUNC00081619 | 4\_2222\_2 | 691 |
| 4\_2222\_2 | JUNC00081618 | 5\_2222\_0 | 730 |
| 5\_2222\_0 | JUNC00081617 | 6\_2222\_2 | 554 |

Table S3: Splice junctions that support the placements of the coding exons for *Tsc1* in *D. yakuba*

Each splice junction matched significantly with the splice site as indicated by high splice junction scores and identical splice sites.

One splice junction contained within one of the coding regions, labeled JUNC00081620, is not contained in Table S4. This is due to the splice junction being on the positive strand and contains a splice junction score of 4, which is very low in comparison to the other splice junctions.

## Supplementary Data

The reconciled gene model for *Tsc1* is available in three different file formats in Supplementary Data:

|  |  |
| --- | --- |
| DyakCAF1\_Tsc1.pep | PEP file containing the peptide sequence of the reconciled gene model |
| DyakCAF1\_Tsc1.fasta | FASTA file containing the transcript sequence of the reconciled gene model |
| DyakCAF1\_Tsc1.gff | GFF file containing coordinates of the coding exons within the Dyak\_CAF1 assembly |

## References

Dabora, S.L., Jozwiak, S., Franz, D.N., Roberts, P.S., Nieto, A., Chung, J., Choy, Y.-S., Reeve, M.P., Thiele, E., Egelhoff, J.C., Kasprzyk-Obara, J., Domanska-Pakiela, D., Kwiatkowski, D.J., 2008. Mutational Analysis in a Cohort of 224 Tuberous Sclerosis Patients Indicates Increased Severity of TSC2, Compared with TSC1, Disease in Multiple Organs [WWW Document]. The American Journal of Human Genetics. URL https://www.sciencedirect.com/science/article/pii/S0002929707624720 (accessed 6.4.20).

Gao, X., Pan, D., 1970. TSC1 and TSC2 tumor suppressors antagonize insulin signaling in cell growth [WWW Document]. Genes & Development. URL http://genesdev.cshlp.org/content/15/11/1383.short (accessed 6.4.20).

Potter, C.J., Huang, H., Xu, T., 2001. Drosophila Tsc1 Functions with Tsc2 to Antagonize Insulin Signaling in Regulating Cell Growth, Cell Proliferation, and Organ Size [WWW Document]. Cell. URL https://www.sciencedirect.com/science/article/pii/S0092867401003336 (accessed 6.4.20).