

***In silico* identification of *Drosophila melanogaster* genes encoding RNA polymerase subunits**

Steven J Marygold^{1§}, Nazif Alic², David S Gilmour³ and Savraj S Grewal⁴

¹FlyBase, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, U.K.

²Institute of Healthy Ageing and the Research Department of Genetics, Evolution, and Environment, University College London, London, U.K.

³Pennsylvania State University, Center for Eukaryotic Gene Regulation, University Park, PA, U.S.A.

⁴Clark H Smith Brain Tumour Centre, Arnie Charbonneau Cancer Institute, & Department of Biochemistry and Molecular Biology, University of Calgary, Alberta, Canada

[§]To whom correspondence should be addressed: sjm41@cam.ac.uk

RNAP	New symbol	CG number	Synonyms	Refs	<i>S. cerevisiae</i> ortholog	<i>H. sapiens</i> ortholog
I, II, III	Polr2E	CG11979	Rpb5		RPB5 [45%]	POLR2E [84%]
	Polr2F	CG1163	Rpl118	1	RPO26 (RPB6) [49%]	POLR2F [63%]
	Polr2H	CG11246	Rpb8		RPB8 [35%]	POLR2H [66%]
	Polr2K	CG34186	Rpb12		RPC10 (RPB12) [49%]	POLR2K [87%]
	Polr2L	CG13628	Rpb10		RPB10 [72%]	POLR2L [91%]
I, III	Polr1C	CG3756	-		RPC40 (AC40) [47%]	POLR1C [54%]
	Polr1D	CG10685	l(2)37Cg		RPC19 (AC19) [40%]	POLR1D [43%]
I	Polr1A	CG10122	Rpl1, RPA1	2	RPA190 [35%]	POLR1A [43%]
	Polr1B	CG4033	Rpl135	3-4	RPA135 [42%]	POLR1B [45%]
	Polr1E	CG18600	-		RPA49 [21%]	POLR1E [22%]
	Polr1F	CG13773	-		RPA43 [18%]	POLR1F [27%]
	-	-	-		RPA34 (A34.5) [n/a]	POLR1G [n/a]
	-	-	-		RPA14 [n/a]	-
	Polr1H	CG13418	Rpl12		RPA12 [36%]	POLR1H [39%]
II	Polr2A	CG1554	Rpl215, Rpb1	5-9	RPO21 (RPB1) [49%]	POLR2A [73%]
	Polr2B	CG3180	Rpl140, wimp	10, 1	RPB2 [58%]	POLR2B [86%]
	Polr2C	CG7885	Rpl33, Rpb3		RPB3 [47%]	POLR2C [69%]
	Polr2D	CG43662	Rpb4	11-12	RPB4 [24%]	POLR2D [77%]
	Polr2G	CG31155	Rpb7		RPB7 [41%]	POLR2G [81%]
	Polr2I	CG3284	Rpl15, Rpb9	13-14	RPB9 [51%]	POLR2I [77%]
	Polr2J	CG6840	Rpb11		RPB11 [50%]	POLR2J/J2/J3 [75%]
	Polr2M	CG3773	Gdn1	15	-	POLR2M [24%]
III	Polr3A	CG17209	Rpc160	16	RPO31 (C160) [45%]	POLR3A [57%]
	Polr3B	CG8344	Rpl1128	4	RET1 (C128) [58%]	POLR3B [69%]
	Polr3C	CG12267	-		RPC82 [17%]	POLR3C [34%]
	Polr3D	CG5147	Rpc53	16	RPC53 [21%]	POLR3D [24%]
	Polr3E	CG10582	Sin		RPC37 [24%]	POLR3E [32%]
	Polr3F	CG5380	-		RPC34 [27%]	POLR3F [50%]
	Polr3G	CG33051	-		RPC31 [23%]	POLR3G/GL [32%]
	Polr3H	CG7339	-		RPC25 [39%]	POLR3H [56%]
	Polr3I	CG4875	Rcp		RPC17 [18%]	CRCP (POLRI) [30%]
	Polr3K	CG33785	-		RPC11 [50%]	POLR3K [70%]
MITO	PolrMT	CG4644	mtRNAPol	17	RPO41 [27%]	POLRMT [37%]

Table 1. Genes encoding RNA polymerase subunits of *Drosophila melanogaster*: RNAP: RNA polymerase to which the subunit encoded by each gene belongs; New symbol: proposed symbol for the *Drosophila* gene; CG number: gene model annotation ID; Synonyms: notable synonyms/previous symbols of the *Drosophila* gene; Refs: reference(s) identifying/characterizing the *Drosophila* gene/protein with respect to its RNAP function: 1) Hamilton *et al.*. 1993, 2)

Knackmuss *et al.*. 1997, 3) Kontermann *et al.*. 1989, 4) Seifarth *et al.*. 1991, 5) Greenleaf *et al.*. 1980, 6) Searles *et al.*. 1982, 7) Greenleaf 1983, 8) Biggs *et al.*. 1985, 9) Jokerst *et al.*. 1989, 10) Falkenburg *et al.*. 1987, 11) Muratoglu *et al.*. 2003, 12) Pankotai *et al.*. 2010, 13) Harrison *et al.*. 1992, 14) Liu *et al.*. 1993, 15) Jishage *et al.*. 2018, 16) Filer *et al.*. 2017, 17) Fernández-Moreno *et al.*. 2009; *S. cerevisiae/H. sapiens* ortholog: the yeast/human ortholog of the *Drosophila* gene, with the percentage amino acid identity between the encoded proteins given in square brackets (highest identity given if multiple orthologs/isoforms). Note that aligning POLR2A/RPB1 sequences without their C-terminal domain repeat regions does not alter their % identity appreciably (data not shown). Yeast and human symbols reflect official nomenclature used at SGD (Cherry *et al.*. 2012) and the HGNC (Braschi *et al.*. 2019), respectively, with popular alternative nomenclature (e.g. Griesenbeck *et al.*. 2017) given in round brackets.

Description

Three highly conserved, multisubunit RNA polymerase (RNAP) enzymes, RNAPs I, II, and III, transcribe the eukaryotic nuclear genome (reviewed by Cramer *et al.*. 2008, Vannini and Cramer 2012, Griesenbeck *et al.*. 2017, Cramer 2019). Each one synthesizes different classes of RNA from DNA templates: RNAP I synthesizes the ribosomal RNA precursor that is processed into most ribosomal RNAs (rRNAs), RNAP II makes messenger RNAs (mRNAs) and a variety of non-coding RNAs, and RNAP III synthesizes short, non-coding RNAs including transfer RNAs (tRNAs), the small 5S rRNA and the U6 small nuclear RNA. Each RNAP contains between 12–17 subunits, ten of which form a structurally conserved catalytic core with additional subunits located on the periphery. Notably, five subunits are shared among all three RNAPs and two others are shared between RNAPs I and III. In contrast to the nuclear RNAPs, a single subunit mitochondrial RNAP transcribes the rRNAs, mRNAs and tRNAs of the mitochondrial genome (Arnold *et al.*. 2012).

While much of what we know about eukaryotic RNAP composition and function comes from studies on yeast and human cells, several *Drosophila melanogaster* (hereafter, *Drosophila*) RNAP subunits have also been isolated and characterized, particularly by Greenleaf, Bautz and colleagues in the 1980s–90s (Greenleaf *et al.*. 1980, Searles *et al.*. 1982, Greenleaf 1983, Biggs *et al.*. 1985, Falkenburg *et al.*. 1987, Jokerst *et al.*. 1989, Kontermann *et al.*. 1989, Seifarth *et al.*. 1991, Hamilton *et al.*. 1993, Liu *et al.*. 1993, Knackmuss *et al.*. 1997). Since the publication of the *Drosophila* genome sequence in 2000, the genes encoding all the subunits of RNAP II (Aoyagi and Wassarman, 2000), several subunits of RNAPs I and III (see supplementary data of Filer *et al.*. 2017 and Martínez Corrales *et al.*. 2020) and the mitochondrial RNAP (Fernández-Moreno *et al.*. 2009) have been identified. Nevertheless, a systematic and complete survey of *Drosophila* RNAP genes is lacking, which has resulted in haphazard nomenclature within the fly literature and FlyBase (flybase.org, Thurmond *et al.*. 2019).

We employed a multi-pronged approach to systematically identify all genes encoding *Drosophila* RNAP subunits (see Methods for details). First, we obtained complete lists of RNAP subunits for yeast (*Saccharomyces cerevisiae*) and humans from recent publications and online resources, and used these to identify the *Drosophila* orthologs. Second, we obtained a list of all *Drosophila* genes annotated with relevant Gene Ontology (GO) terms. Importantly, these annotations include those based on direct experimental evidence as well as inferences based on sequence similarity/orthology and the presence of defined protein domains. Finally, we searched the *Drosophila* literature for reports of individual, or lists of, RNAP subunits. The results of these three approaches were cross-checked and integrated, and the results are presented in Table 1.

We find that a total of 31 distinct genes encode RNAP subunits in *Drosophila*. We identified genes encoding the five subunits shared between RNAPs I, II and III as well as the two subunits shared by RNAPs I and III. We also identified genes encoding an additional five subunits of RNAP I, an additional eight subunits of RNAP II, an additional ten subunits of RNAP III and the mitochondrial RNAP. Thus, *Drosophila* possesses twelve RNAP I subunits, thirteen RNAP II subunits, seventeen RNAP III subunits, and a single mitochondrial RNAP. Only a third of these have been characterized directly in *Drosophila*, either biochemically or genetically, with research having focussed on RNAP II subunits and the largest subunits of RNAPs I and III (see Refs column of Table 1). The *Drosophila* subunits show a range of 17–72% (mean of 39%) and 22–91% (mean of 55%) amino acid identity to their orthologs in *S. cerevisiae* and humans, respectively. A comparison of the complement of RNAP subunits across those three species reveals four notable differences: (i) *Drosophila* lacks an identifiable ortholog of yeast RPA34/human POLR1G (Martínez Corrales *et al.*. 2020); (ii) neither *Drosophila* or humans have an ortholog of yeast RPA14 (Russell and Zomerdijk 2006; Martínez Corrales *et al.*. 2020); (iii) yeast lack the POLR2M subunit, which defines a metazoan-specific RNAP II subpopulation (Hu *et al.*. 2006); and (iv) humans possess multiple copies of genes encoding RPB11/POLR2J and RPB7/POLR3G, whereas these are single-copy genes in *Drosophila* and yeast.

Prior to this study, 22 of the 31 *Drosophila* RNAP genes had been named in FlyBase using a variety of conventions. Seven were named based on the empirically determined molecular weight of the *Drosophila* proteins (*RpII18*, *RpI1*, *RpI135*, *RpII215*, *RpII140*, *RpIII15*, *RpIII128*), following a nomenclature originally proposed in Greenleaf *et al.*. 1980. Fourteen RNAP genes had been named after their yeast or human ortholog, and one additional gene (*Sin*) was named for an unrelated physical interaction (Dong and Bell, 1999). The remaining nine genes were unnamed or had only a ‘placeholder’ symbol. We wished to assign an informative, systematic nomenclature to all *Drosophila* RNAP genes.

Unfortunately, a universal eukaryotic RNAP nomenclature system does not exist, with two different systems currently in use for yeast and humans/vertebrates (Table 1). We propose that the human nomenclature system is adopted for the *Drosophila* genes in FlyBase for the following reasons: (i) individual *Drosophila* RNAP subunits show greater identity to the human subunits compared to yeast; (ii) the overall complement of *Drosophila* RNAP subunits is more similar to humans than yeast; (iii) unlike the yeast nomenclature, the human nomenclature follows a systematic format for all subunits; (iv) using the human nomenclature for the *Drosophila* subunits will facilitate the use/comparison of *Drosophila* data in biomedicine. (The yeast nomenclature will be retained/added to the *Drosophila* gene reports as searchable and browsable synonyms.)

In conclusion, our complete and rationalized listing of *Drosophila* RNAP subunits will be useful to *Drosophila* researchers working in this field as well as to those wishing to compare RNAP biology between fly, yeast, human and other species.

Methods

[Request a detailed protocol](#)

Publications identifying/characterizing *Drosophila* RNAP subunits were identified using PubMed (pubmed.ncbi.nlm.nih.gov), FlyBase (flybase.org, Thurmond *et al.* 2019) and Google (www.google.com). Published lists of *S. cerevisiae* and human RNAP subunits were obtained from Huang and Maraia 2001, Hu *et al.* 2002, Russell and Zomerdijk 2006, Cramer *et al.* 2008, Vannini and Cramer 2012 and Griesenbeck *et al.* 2017. In addition, a curated list of human RNAP subunits was obtained from the HGNC (www.genenames.org/data/genegroup/#!/group/726, Braschi *et al.* 2019). Ortholog predictions and protein identity percentages were obtained from the integrative ortholog prediction tool, DIOPT (v8) (Hu *et al.* 2011) via FlyBase. All reported orthologs in Table 1 are reciprocal best hits, with the exception of human POLR3G and POLR3G paralogs, where all genes are listed. Orthology predictions were verified using the HCOP tool (Eyre *et al.* 2007). The Alliance of Genome Resources database (www.alliancegenome.org (release 3.1.1), The Alliance of Genome Resources Consortium, 2020) was used to query fly, yeast and human for relevant GO annotations (using terms RNA polymerase I complex (GO:0005736), RNA polymerase II, core complex (GO:0005665), RNA polymerase III complex (GO:0005666) and mitochondrial DNA-directed RNA polymerase complex (GO:0034245)). Gene symbol information was obtained from FlyBase (FB2020_03), SGD (www.yeastgenome.org, accessed 17th August 2020) and HGNC (www.genenames.org, accessed 17th August 2020).

Acknowledgments: We thank Kevin Cook, Julia Zeitlinger and Joan Conaway for comments on the manuscript, and Elspeth Bruford and Bryony Braschi at the HGNC for discussions on human RNAP gene nomenclature.

References

- Aoyagi N, Wassarman DA. 2000. Genes encoding *Drosophila melanogaster* RNA polymerase II general transcription factors: diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. *J Cell Biol* 150: F45-50. PMID: 10908585.
- Arnold JJ, Smidansky ED, Moustafa IM, Cameron CE. 2012. Human mitochondrial RNA polymerase: structure-function, mechanism and inhibition. *Biochim Biophys Acta* 1819: 948-60. PMID: 22551784.
- Biggs J, Searles LL, Greenleaf AL. 1985. Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of *Drosophila* RNA polymerase II. *Cell* 42: 611-21. PMID: 2992806.
- Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, Yates B, Bruford E. 2019. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res* 47: D786-D792. PMID: 30304474.
- Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED. 2012. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40: D700-5. PMID: 22110037.
- Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, Jawhari A, Jennebach S, Kamenski T, Kettenberger H, Kuhn CD, Lehmann E, Leike K, Sydow JF, Vannini A. 2008. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* 37: 337-52. PMID: 18573085.
- Cramer P. 2019. Eukaryotic Transcription Turns 50. *Cell* 179: 808-812. PMID: 31675494.
- Dong Z, Bell LR. 1999. SIN, a novel *Drosophila* protein that associates with the RNA binding protein sex-lethal. *Gene* 237: 421-8. PMID: 10521666.
- Eyre TA, Wright MW, Lush MJ, Bruford EA. 2007. HCOP: a searchable database of human orthology predictions. *Brief Bioinform* 8: 2-5. PMID: 16951416.

- Falkenburg D, Dworniczak B, Faust DM, Bautz EK. 1987. RNA polymerase II of *Drosophila*. Relation of its 140,000 Mr subunit to the beta subunit of *Escherichia coli* RNA polymerase. *J Mol Biol* 195: 929-37. PMID: 3116266.
- Fernández-Moreno MA, Bruni F, Adán C, Sierra RH, Polosa PL, Cantatore P, Garesse R, Roberti M. 2009. The *Drosophila* nuclear factor DREF positively regulates the expression of the mitochondrial transcription termination factor DmTTF. *Biochem J* 418: 453-62. PMID: 19032147.
- Filer D, Thompson MA, Takhaveev V, Dobson AJ, Kotronaki I, Green JWM, Heinemann M, Tullet JMA, Alic N. 2017. RNA polymerase III limits longevity downstream of TORC1. *Nature* 552: 263-267. PMID: 29186112.
- Greenleaf AL, Weeks JR, Voelker RA, Ohnishi S, Dickson B. 1980. Genetic and biochemical characterization of mutants at an RNA polymerase II locus in *D. melanogaster*. *Cell* 21: 785-92. PMID: 6777048.
- Greenleaf AL. 1983. Amanitin-resistant RNA polymerase II mutations are in the enzyme's largest subunit. *J Biol Chem* 258: 13403-6. PMID: 6417125.
- Griesenbeck J, Tschochner H, Grohmann D. 2017. Structure and Function of RNA Polymerases and the Transcription Machineries. *Subcell Biochem* 83: 225-270. PMID: 28271479.
- Hamilton BJ, Mortin MA, Greenleaf AL. 1993. Reverse genetics of *Drosophila* RNA polymerase II: identification and characterization of RpII140, the genomic locus for the second-largest subunit. *Genetics* 134: 517-29. PMID: 8325487.
- Harrison DA, Mortin MA, Corces VG. 1992. The RNA polymerase II 15-kilodalton subunit is essential for viability in *Drosophila melanogaster*. *Mol Cell Biol* 12: 928-35. PMID: 1545824.
- Hu P, Wu S, Sun Y, Yuan CC, Kobayashi R, Myers MP, Hernandez N. 2002. Characterization of human RNA polymerase III identifies orthologues for *Saccharomyces cerevisiae* RNA polymerase III subunits. *Mol Cell Biol* 22: 8044-55. PMID: 12391170.
- Hu X, Malik S, Negroiu CC, Hubbard K, Velalar CN, Hampton B, Grosu D, Catalano J, Roeder RG, Gnatt A. 2006. A Mediator-responsive form of metazoan RNA polymerase II. *Proc Natl Acad Sci U S A* 103: 9506-11. PMID: 16769904.
- Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE. 2011. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* 12: 357. PMID: 21880147.
- Huang Y, Maraia RJ. 2001. Comparison of the RNA polymerase III transcription machinery in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and human. *Nucleic Acids Res* 29: 2675-90. PMID: 11433012.
- Jishage M, Yu X, Shi Y, Ganesan SJ, Chen WY, Sali A, Chait BT, Asturias FJ, Roeder RG. 2018. Architecture of Pol II(G) and molecular mechanism of transcription regulation by Gdown1. *Nat Struct Mol Biol* 25: 859-867. PMID: 30190596.
- Jokerst RS, Weeks JR, Zehring WA, Greenleaf AL. 1989. Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*. *Mol Gen Genet* 215: 266-75. PMID: 2496296.
- Knackmuss S, Bautz EF, Petersen G. 1997. Identification of the gene coding for the largest subunit of RNA polymerase I (A) of *Drosophila melanogaster*. *Mol Gen Genet* 253: 529-34. PMID: 9065685.
- Kontermann R, Sitzler S, Seifarth W, Petersen G, Bautz EK. 1989. Primary structure and functional aspects of the gene coding for the second-largest subunit of RNA polymerase III of *Drosophila*. *Mol Gen Genet* 219: 373-80. PMID: 2482932.
- Liu Z, Kontermann RE, Schulze RA, Petersen G, Bautz EK. 1993. RPII15 codes for the M(r) 15,000 subunit 9 of *Drosophila melanogaster* RNA polymerase II. *FEBS Lett* 335: 73-5. PMID: 8243669.
- Martínez Corrales G, Filer D, Wenz KC, Rogan A, Phillips G, Li M, Feseha Y, Broughton SJ, Alic N. 2020. Partial Inhibition of RNA Polymerase I Promotes Animal Health and Longevity. *Cell Rep* 30: 1661-1669.e4. PMID: 32049000.
- Muratoglu S, Georgieva S, Pápai G, Scheer E, Enünlü I, Komonyi O, Cserpán I, Lebedeva L, Nabirochkina E, Udvardy A, Tora L, Boros I. 2003. Two different *Drosophila* ADA2 homologues are present in distinct GCN5 histone acetyltransferase-containing complexes. *Mol Cell Biol* 23: 306-21. PMID: 12482983.
- Pankotai T, Ujfaludi Z, Vámos E, Suri K, Boros IM. 2010. The dissociable RPB4 subunit of RNA Pol II has vital functions in *Drosophila*. *Mol Genet Genomics* 283: 89-97. PMID: 19921261.
- Russell J, Zomerdijk JC. 2006. The RNA polymerase I transcription machinery. *Biochem Soc Symp* : 203-16. PMID: 16626300.
- Searles LL, Jokerst RS, Bingham PM, Voelker RA, Greenleaf AL. 1982. Molecular cloning of sequences from a *Drosophila* RNA polymerase II locus by P element transposon tagging. *Cell* 31: 585-92. PMID: 6297774.
- Seifarth W, Petersen G, Kontermann R, Riva M, Huet J, Bautz EK. 1991. Identification of the genes coding for the second-largest subunits of RNA polymerases I and III of *Drosophila melanogaster*. *Mol Gen Genet* 228: 424-32. PMID: 1910149.

The Alliance of Genome Resources Consortium. 2020. Alliance of Genome Resources Portal: unified model organism research platform. *Nucleic Acids Res* 48: D650-D658. PMID: 31552413.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, Kaufman TC, Calvi BR, FlyBase Consortium. 2019. FlyBase 2.0: the next generation. *Nucleic Acids Res* 47: D759-D765. PMID: 30364959.

Vannini A, Cramer P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* 45: 439-46. PMID: 22365827.

Funding: S.J.M. is funded by a grant from the National Human Genome Research Institute of the NIH [U41HG000739] to Norbert Perrimon (PI), Nicholas Brown (co-PI). N.A. is funded by grants from the BBSRC [BB/S014357/1 and BB/R014507/1], D.S.G. is funded by a grant from the National Institute of General Medical Sciences of the NIH [R01-GM0474777], and S.S.G is funded by a project grant from the Canadian Institutes of Health Research.

Author Contributions: Steven J Marygold: Conceptualization, Writing - original draft, Methodology, Investigation, Data curation. Nazif Alic: Validation, Writing - review and editing. David S Gilmour: Validation, Writing - review and editing. Savraj S Grewal: Validation, Writing - review and editing.

Reviewed By: Anonymous

History: **Received** September 21, 2020 **Revision received** October 12, 2020 **Accepted** October 16, 2020 **Published** October 20, 2020

Copyright: © 2020 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Marygold, SJ; Alic, N; Gilmour, DS; Grewal, SS (2020). *In silico* identification of *Drosophila melanogaster* genes encoding RNA polymerase subunits. *microPublication Biology*. <https://doi.org/10.17912/micropub.biology.000320>