


SUPPLEMENTARY MATERIAL

corresponding to:

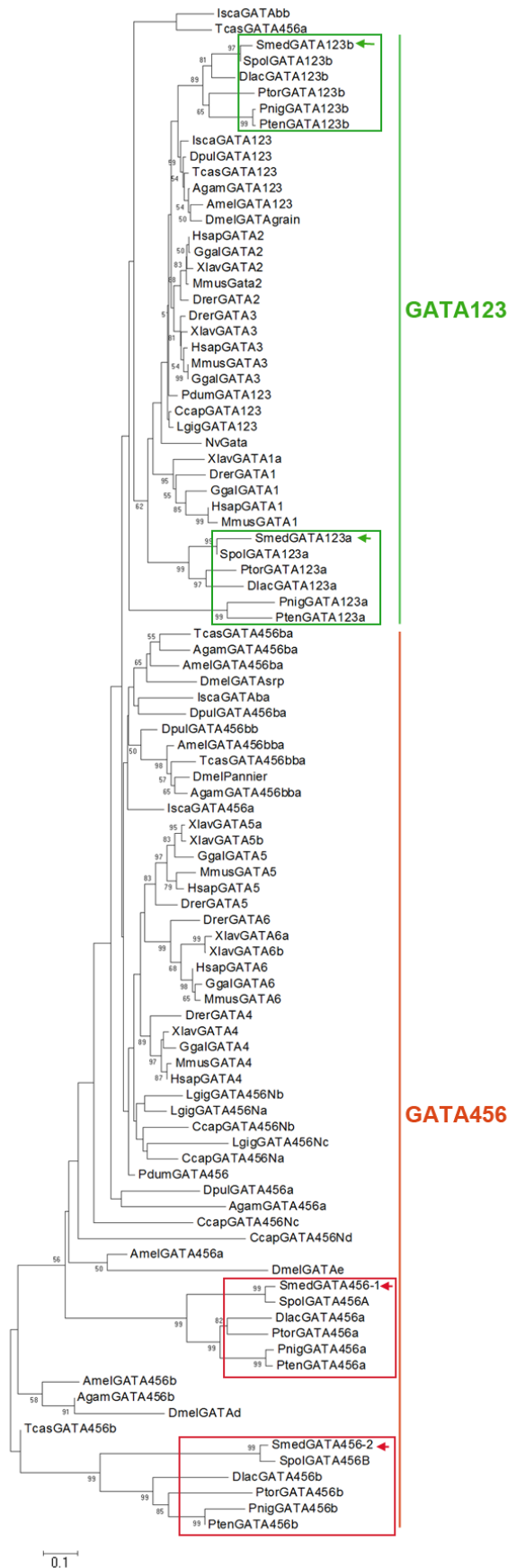
**The pioneer factor *Smed-gata456-1* is required for
gut cell differentiation and maintenance in planarians**

ALEJANDRO GONZÁLEZ-SASTRE, NÍDIA DE SOUSA, TERESA ADELL and EMILI SALÓ*

*Address correspondence to: Emili Saló. Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Av. Diagonal 643, edifici annex planta 1, 08028 Barcelona, Catalunya, Spain. Tel: +34-934035977. Fax +34-934034420. E-mail: esalo@ub.edu - web: <http://planarian.bio.ub.es> -  <http://orcid.org/0000-0001-5460-9223>

Suppl. Fig. 1. Phylogenetic analysis of GATA transcription factors.

GATA sequences from planarian species were identified in the Planmine database (Brandl et al., 2016) and added to those used in the phylogenetic analysis of Gilles et al., (2008). Planarian sequences are indicated with a square. Schmidtea mediterranea sequences are marked with an arrow. GATA123 and GATA456 family members were analysed. Abbreviations: Agam, Anopheles gambiae; Amel, Apis melifera; Ccap, Capitella capitata; Dlac, Dendrocoelum lacteum; Dmel, Drosophila melanogaster; Dpul, Daphnia pulex; Drer, Danio rerio; Ggal, Gallus gallus; Hsap, Homo sapiens; Isca, Ixodes scapularis; Lgig, Lottia gigantea; Mmus, Mus musculus; Nv, Nematostella vectensis; Pdum, Platynereis dumerilii; Pnig, Polycelis nigra; Pten, Polycelis tenuis; Ptor, Planaria torva; Smed, Schmidtea mediterranea; Spol, Schmidtea polychroa; Tcas, Tribolium castaneum; Xlav, Xenopus laevis.

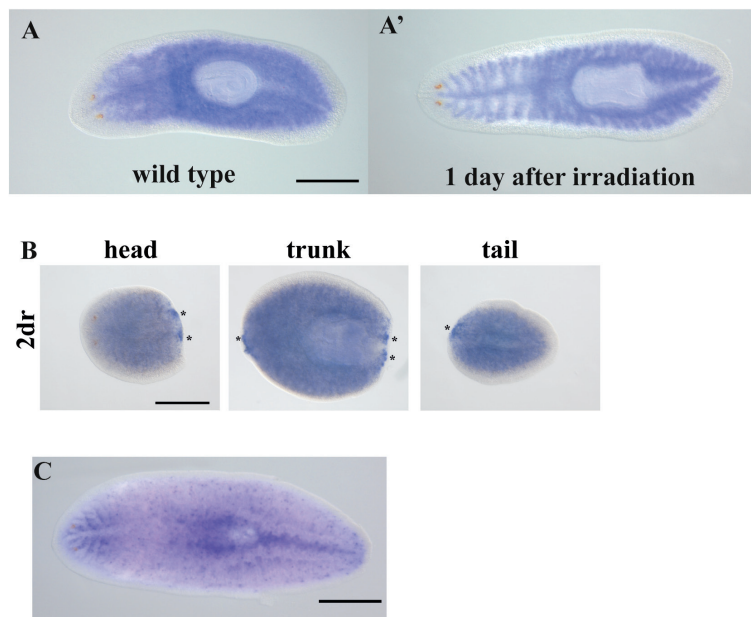


MmusGATA5	PQ--KRL---	-----	-----	-----SSSR	RAGL	CSNCH	TATTTLWRRN	SEGEPCVNAC
XlavGATA5b	PQ--KRL---	-----	-----	-----SSSR	RAGL	CTNCH	TSSTTLWRRN	SEGEPCVNAC
XlavGATA5a	PQ--KRL---	-----	-----	-----SSSR	RAGL	CTNCH	TSSTTLWRRN	SEGEPCVNAC
GgalGATA5	PQ--KRL---	-----	-----	-----SSSR	RAGL	CTNCH	TTNTTLWRRN	AEGEPCVNAC
DrerGATA5	PQ--KRL--Q	-----	-----	-----STSR	RAGL	CTNCH	TSSTTLWRRN	AEGEPCVNAC
CcapGATA456Nb	PQ--KRM---	-----	-----	-----CASR	RMGL	QCANCQ	TSSTTLWRRN	ADGEPCVNAC
DpulGATA456bb	PP--RRL--L	-----	-----	-----SATR	RLGL	CTNCG	TTTTTLWRRN	AEGEPCVNAC
DmelGATAsrp	QP--RRL---	-----	-----	-----SASK	RAGL	CSNCH	TTHTSLWRRN	PAGEPCVNAC
AgamGATA456ba	QP--RRLVKD	PVNQSF---	-----	-----NYPHQSSAR	RVGL	QCNCN	TTNTSLWRRN	QVGEPCVNAC
DpulGATA456ba	QT--RRL---	-----	-----	-----STTR	RLGL	RCANCA	TTTTSLWRRN	NQGETPCVNAC
TcasGATA456a	PP-----	-----	-----	-----KK	RNGV	QCANCQ	TGNTTLWRRN	NQGETPCVNAC
DlacGATA123b	PK--RRL---	-----	-----	-----SASR	RTGT	CSNCN	TSSTTLWRRN	TNGEPCVNAC
DpulGATA456a	QP--PS---	-----	-----	-----SGNR	RAGL	CSNCN	TSSTTLWRRN	ANGEPCVNAC
LgigGATA456Nc	PN--LTE---	----NR---	-----	-----TVSLGFSSR	RMGL	ACANCM	TSSTTLWRRN	AEGEPCVNAC
TcasGATA456bb	-----	-----	-----	-----AR	RVDL	CTNCG	TQTTIWRNR	MKGEMVCNAC
AmelGATA456bb	-----	-----	-----	-----AK	KVDM	CTNCG	TMTTIIWRNR	MKGEMVCNAC
AgamGATA456bb	-----	-----	-----	-----NSQ	QKDM	CTNCG	TTTTTIWRNR	IRGEMVCNAC
DmelGATAd	-----	-----	-----	-----	--DM	CSNCG	TLTTTIWRNR	VRGEMVCNAC
CcapGATA456Nc	TSGGRRV---	-----	-----	-----SVSR	RVGL	SCANCH	TSSTTLWRRN	NEGEPCVNAC
AgamGATA456a	SQ--KAKQT-	---VVS---	-----	-----GTPPAQGNR	RSGV	TCANCQ	TTTTTLWRRN	NQGDPCVNAC
IscaGATAbb	PP-----	-----	-----	-----	-----	-----	-----	-----
PnigGATA123a	PK--RRL---	-----	-----	-----STR	KMGT	BCINCG	TTHTTLWRRN	QQGDSVCNAC
PtenGATA123a	QK--QKL---	-----	-----	-----SR	KMGA	SCINCG	TTHTTLWRRN	QQGDSVCNAC
AmelGATA456a	-----	-----	-----	-----R	RTGV	DCANCR	TSNTTLWRRN	NNGEPCVNAC
CcapGATA456Nd	PL--RMA---	-----	-----	-----SSSS	RETH	CANCQ	TQQTMMWRNR	TDGEPCVNAC
SmedGATA456-2	KQ-----	-----	-----	-----TAK	RTGL	CSNCK	TENTTLWRRN	SEGQPCVNAC
SpolGATA456B	KQ-----	-----	-----	-----TAK	RTGL	CSNCK	TENTTLWRRN	SEGQPCVNAC
DlacGATA456b	KQ-----	-----	-----	-----TAK	RTGL	CSNCK	SDHTTLWRRN	AEGQPCVNAC
PnigGATA456b	KQ-----	-----	-----	-----TAK	RTGL	CSNCK	SDHTTLWRRN	AEGLPVCNAC
PtenGATA456b	KQ-----	-----	-----	-----TAK	RTGL	CSNCK	SDHTTLWRRN	AEGLPVCNAC
PtorGATA456b	KQ-----	-----	-----	-----TAK	RSGL	CSNCK	SDHTTLWRRN	AEGQPCVNAC
DmelGATAe	PK--AAA---	-----	-----	-----APNNR	RNGV	TCANCQ	TNSTTLWRRN	NEGNPCVNAC

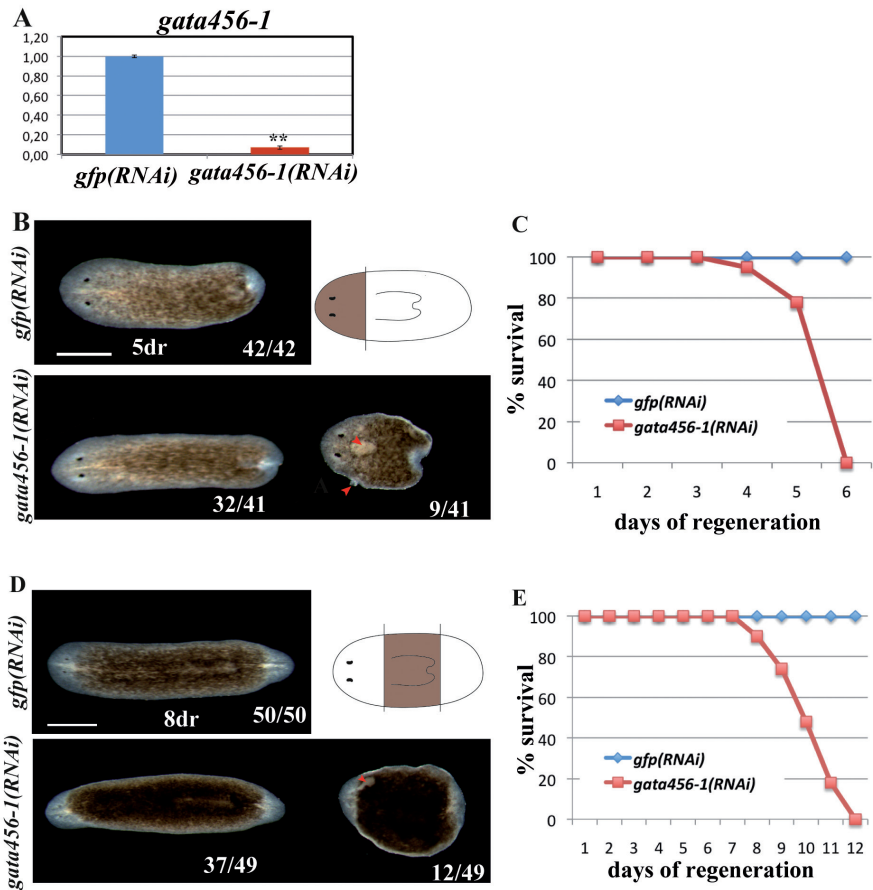
		150	160	170				
SmedGATA456-1	GLYYKLHKS	RPLSMRKEGI	QTRKRKRKTQ	MKPVSTH				
SpolGATA456A	GLYYKLHKS	RPLSMRKEGI	QTRRRRKRKTQ	MKPVSTH				
DlacGATA456a	GLYYKLHKS	RPLSMRKEGI	QTRKRKRKQ	MKPVSTH				
PtorGATA456a	GLYYKLHKS	RPLSMRKEGI	QTRKRKRKSQ	LKPVNSQ				
PnigGATA456a	GLYYKLHKS	RPLSMRKEGI	QTRKRKRKSQ	LKPVSSQ				
PtenGATA456a	GLYYKLHKS	RPLSMRKEGI	QTRKRKRKSQ	LKPVNSQ				
SmedGATA123b	GLYFKLHSV	RPPTMKKEGI	QTRNRKVSHK	SKRRRSE				
SpolGATA123b	GLYFKLHSV	RPPTMKKEGI	QTRNRKVSHK	SKRRRSE				
SmedGATA123a	GLYYKLHHIS	RPISMKKDLI	QTRNRKITKS	KKKKEFD				
Spol_GATA123a	GLYYKLHHIS	RPISM---	-----	-----				
PtorGATA123a	GLYYKLHHIN	RPISMKKDLI	QTRNRKITKS	KKKRDFF				
DlacGATA123a	GLYYKLHHIN	RPISMKKDLI	QTRNRKITKS	KKKRDFF				
IscaGATA123	GLYFKLHNVN	RPLTMKKEGI	QTRNRKLSSK	SKKKKGL				
DpulGATA123	GLYYKLHNVN	RPLTMKKEGI	QTRNRKLSSK	SKKKKGL				
TcasGATA123	GLYYKLHNVN	RPLTMKKEGI	QTRNRKLSSK	SKKKKSG				
AgamGATA123	GLYYKLHNVN	RPLTMKKEGI	QTRNRKLSSK	SKKKKGI				
AmelGATA123	GLYYKLHNVN	RPLTMKKEGI	QTRNRKLSSK	SKKKKAG				
DmelGATAgrain	GLYYKLHNVN	RPLTMKKEGI	QTRNRKLSSK	SKKKKGL				
HsapGATA2	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKSKK				
MmusGata2	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKSKK				
GgalGATA2	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKSK--				
XlavGATA2	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKSK--				
DrerGATA2	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKSK--				
HsapGATA3	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKCKKV				
MmusGATA3	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKCK--				
GgalGATA3	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKCK--				
XlavGATA3	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKSK--				
DrerGATA3	GLYYKLHNVN	RPLTMKKEGI	QTRNRKMSNK	SKKSK--				
CcapGATA123	GLYYKLHNVN	RPLTMKDDGI	QTPNPKMS--	-----				
PdumGATA123	GLYYKLHNVN	RPLTMKDDGI	QTRNRKMSNK	SKKSKK				
LgigGATA123	GLYYKLHNVN	RPLTMKDDGI	QTRNRKMSNK	SKKSKK				
NvGata	GLYKLVHNVN	RPLSMKDDGI	QTRNRKMSNK	SKKSKK				
HsapGATA1	GLYYKLHNVN	RPLTMKDDGI	QTRNRKMSNK	GKKKRG				
MmusGATA1	GLYKLVHNVN	RPLTMKDDGI	QTRNRKMSNK	GKKKRG				
GgalGATA1	GLYYKLHNVN	RPLTMKDDGI	QTRNRKMSNK	GKKKRP				
XlavGATA1a	GLYYKLHNVN	RPLTMKDDGI	QTRNRKMSNK	SKKSK--				
DrerGATA1	GLYKLVHNVN	RPLTMKDDGI	QTRNRKMSNK	NKKK--				
PtorGATA123b_	GLYKLVHNVN	RPPTMKDDGI	QTRNRKMSNK	SKRRRSE				
PnigGATA123b_	GLYKLVHNVN	RPPTMKDDGI	QTRNRKMSNK	SKRRRSE				
PtenGATA123b_	GLYKLVHNVN	RPPTMKDDGI	QTRNRKMSNK	SKRRRSE				
IscaGATAba	GLYKLVHNVN	RPLAMKDDSI	QTRKRKPKNS	SSGSGS				
TcasGATA456ba	GLYKLVHNVN	RPLAMKDDSI	QTRKRKPKNS	KDSNSRN				
AmelGATA456ba	GLYKLVHNVN	RPLAMKDDSI	QTRKRKPKNS	MKSSDTP				
IscaGATA456a	GLYKLVHNVN	-----	-----	-----				
TcasGATA456bb	GLYKLVHNVN	RPLAMKDDGI	QTRKRKMSNK	AAAAAAV				
AmelGATA456bb	GLYKLVHNVN	RPLAMKDDGI	QTRKRKPKNS	PEPNARS				
DmelPannier	GLYKLVHNVN	RPLAMKDDGI	QTRKRKPKNS	GGSGAVG				
AgamGATA456bb	GLYKLVHNVN	RPLAMKDDGI	QTRKRKPKNS	GGSGGSA				

LgigGATA456Na	GLYYKHLGVH	RPLAMKDDGI	QTRKRKPKSK	ERSPHKT
PdumGATA456	GLYYKHLGVN	RPLAMKKEGI	QTKKKK----	-----
HsapGATA4	GLYMKLHGVP	RPLAMRKEGI	QTRKRKPKNL	NKSKTPA
MmusGATA4	GLYMKLHGVP	RPLAMRKEGI	QTRKRKPKNL	NKSKTPA
GgalGATA4	GLYMKLHGVP	RPLAMRKEGI	QTRKRKPKNL	NKTKTP-
XlavGATA4	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNL	SKSKTLT
DrerGATA4	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	SKTKPGS
HsapGATA6	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	NKSKTCS
MmusGATA6	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	NKSKACS
GgalGATA6	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	NKSKAC-
XlavGATA6b	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNL	NKSKSSS
XlavGATA6a	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKTL	NKSKSSS
DrerGATA6	GLYTKLHGVP	RPLAMKKEGI	QTRKRKPKSL	SKVKGSS
CcapGATA456Na	GLYYKLHQVN	RPISMKDDGI	QTRKRKPKGS	GKSKSKH
LgigGATA456Nb	GLYYKHLGVN	RPLAMKDDGI	QTRKRKPKQG	GKTKSSS
HsapGATA5	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKTI	AKARGSS
MmusGATA5	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	AKIKGSS
XlavGATA5b	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	KGKGTST
XlavGATA5a	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	KGKGTST
GgalGATA5	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	TKGKTST
DrerGATA5	GLYMKLHGVP	RPLAMKKEGI	QTRKRKPKNI	TKGKTST
CcapGATA456Nb	GLYFKLHGVP	RPMSMRKDDGI	QTRKRKPKNL	NKGPSK-
DpulGATA456bb	GLYHKLHGVP	RPLAMRKEGI	QTRKRKPKSN	SA-----
DmelGATA5srp	GLYYKLHNSV	RPLTMKDDTI	QTRKRKPKGT	KSEKSKS
AgamGATA456ba	GLYYKLHNVN	RPLAMKDDNI	QSRKRKPKGS	KNSDGNG
DpulGATA456ba	GLYFKLHGVP	RPLAMKDDNI	QTRKRKPKGD	TFVNVK
TcasGATA456a	GLYFKLHNVN	RPLSMKKEGI	QTRKRKPKSS	NHSQPA
DlacGATA123b	GLYFKLHNSV	RPP-----	-----	-----
DpulGATA456a	GLYFKLHNVC	RPLAMKKEGI	QTRKRKPKAG	TPSGAPT
LgigGATA456Nc	GLYYKLHQVN	RPMSMKDDGI	QTRKRKPKTP	SKLKTSP
TcasGATA456bb	GLYYKHLGID	RPHTMRRDTI	HTRRRR----	-----
AmelGATA456bb	GLYYKHLGVN	RPVTMRRDTI	HTRRRRPKGE	KPTRHRK
AgamGATA456bb	GLYFKLHGVP	RPHTMRRDTI	HTRRRR----	-----
DmelGATA4d	GLYFKLHGVP	RPHTMRRDTI	HTRRRRPKEL	ERSKSKH
CcapGATA456Nc	GLYFKLHGVP	RPMSMKKEGI	QTRKRKPKGS	GKQKSSP
AgamGATA456a	GLYYKLHNSV	RPLTMKDDGI	QTRKRKPKSS	QQIQPMN
IscaGATA4bb	-----K	RPV-----	-----	-----
PnigGATA123a	GLYFKLHHIH	RPMSMKKEGI	QTRHRKVNKF	KKSYPP
PtenGATA123a	GLYFKLHHVH	RPMSMKKEGI	QTRRRKATKY	KSHPFKH
AmelGATA456a	GLYYKLHNVN	RPLSMKKEGI	QTRKRKPKNN	SGISGNL
CcapGATA456Nd	GLYFKLHKVN	RPLSMVKEGI	QSRKRKPKSS	SSPGIMP
SmedGATA456-2	GLYYRLHKTH	RPPTMRKEIL	QSRKRKTKKK	QTMCMKY
SpolGATA456B	GLYYRLHKTH	RPPTMRKEIL	QSRKRKTKKK	QTMCMKY
DlacGATA456b	GLYYRLHKTA	RPPSMRKETL	QTRKRKTKRK	PSLCMKY
PnigGATA456b	GLYYRLHKTA	RPPSMRKETL	QTRKRKTKRK	PSICMKY
PtenGATA456b	GLYYRLHKTA	RPPSMRKETL	QTRKRKTKRK	PNICMKY
PtorGATA456b	GLYFRLHKTA	RPPSMRKETL	QTRKRKTKRK	PSFGMKY
DmelGATAe	GLYYKLHNMN	RPLSMKKEGI	QKRKRKPKNN	GG-----

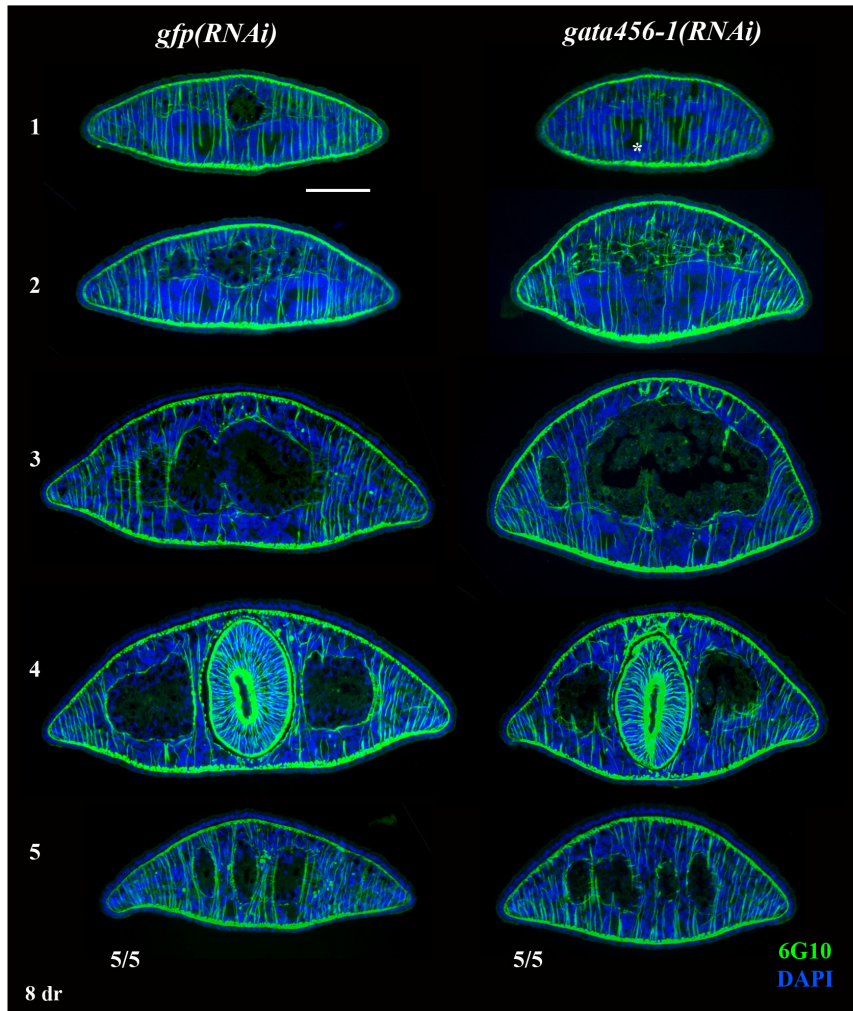
Suppl. Fig. 2. Alignment of the conserved dual zinc-finger domain of GATA transcription factors corresponding to the phylogenetic analysis shown in Sup Fig. 1. Zn-finger motives are boxed. Abbreviations are indicated in Supp. Fig. 1.



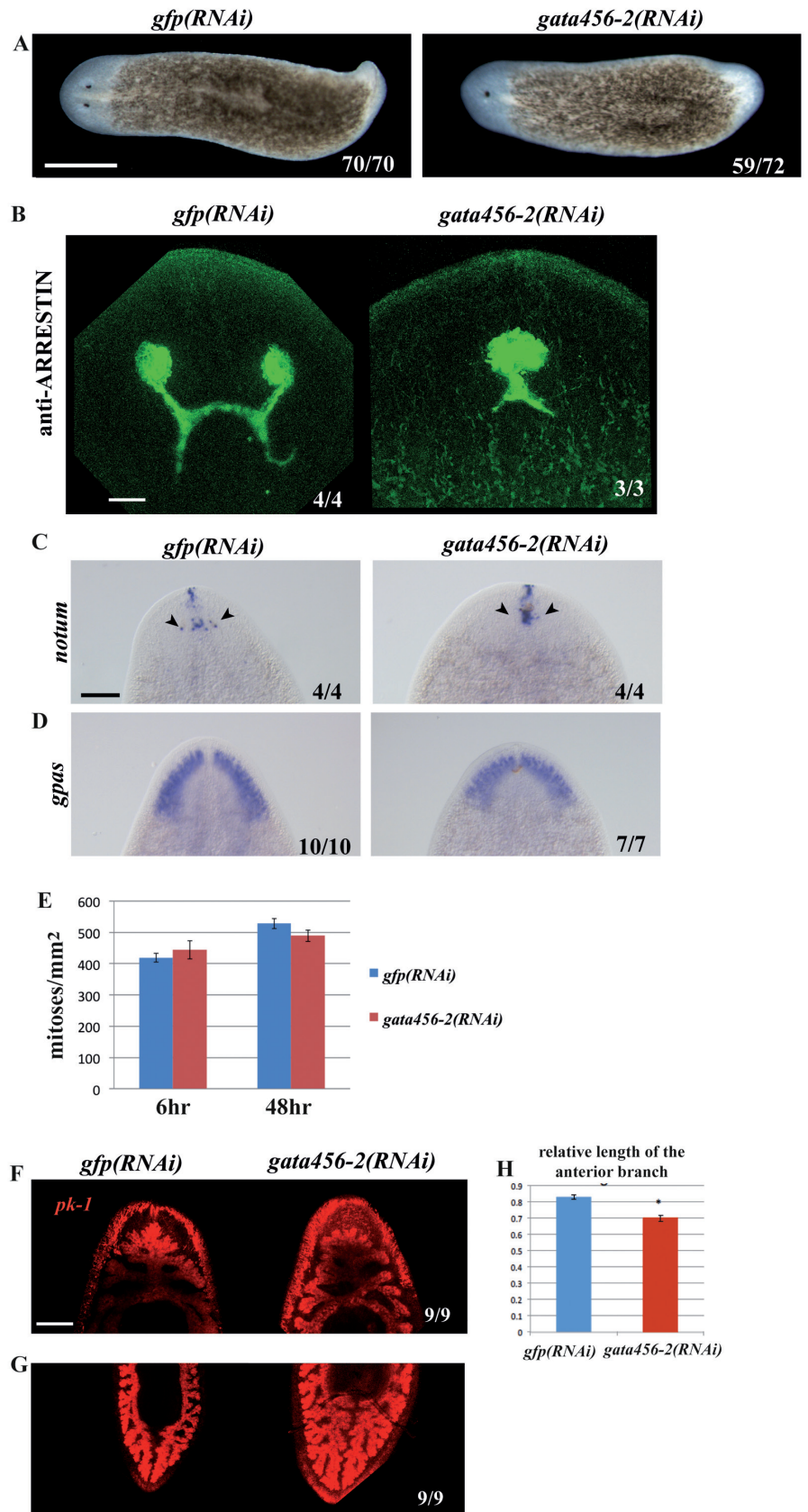
Suppl. Fig. 3. Expression patterns of *Smed-gata456-1* and *Smed-gata456-2*. (A) Expression pattern of *Smed-gata456-1* in intact wild type animals and 1 day after irradiation, as determined by whole mount in situ hybridization. (B) Expression pattern of *Smed-gata456-1* at 2 days of regeneration as determined by whole mount in situ hybridization. Asterisks indicate regenerating gut branches. (C) Expression pattern of *Smed-gata456-2* in intact animals as determined by whole mount in situ hybridization. The anterior end is oriented to the left. dr, days of regeneration. Scale bar: 500 μm .



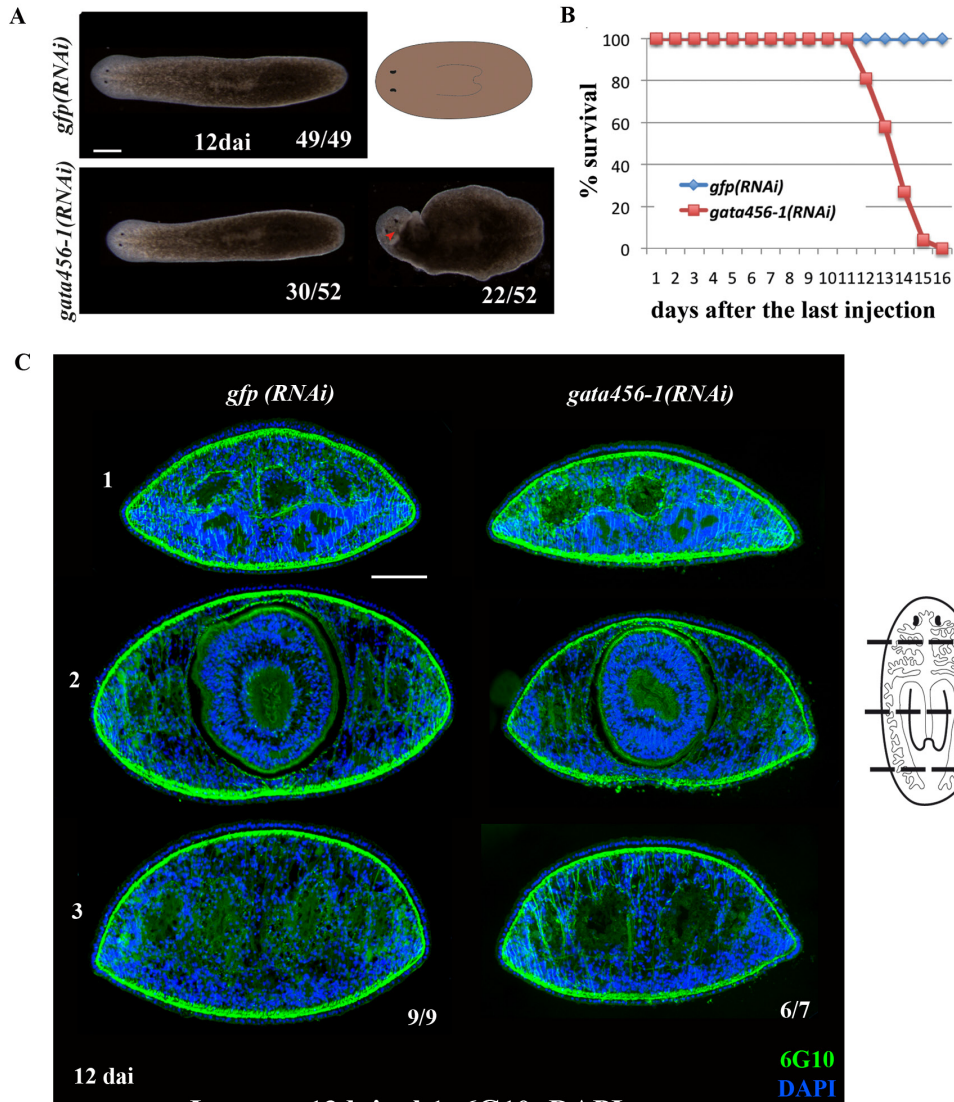
Suppl. Fig. 4. *Smed-gata456-1(RNAi)* is lethal in regenerating animals. (A) Quantitative real time PCR (qRT-PCR) for *Smed-gata456-1* in control and RNAi animals after 8 days of regeneration. Error bars represent the standard error of the mean. Data were analyzed by Student's *t*-test. ** $P < 0.001$. (B) Live regenerating control and *Smed-gata456-1(RNAi)* heads after 5 days of regeneration. (C) Graph showing survival of *Smed-gata456-1(RNAi)* regenerating heads ($n=41$). (D) Live regenerating control and *Smed-gata456-1(RNAi)* trunks after 8 days of regeneration. (E) Graph showing survival of *Smed-gata456-1(RNAi)* regenerating trunks ($n=49$). Red arrowheads indicate lesions. dr, days of regeneration. The anterior end is oriented to the left. Scale bar: 500 μ m.



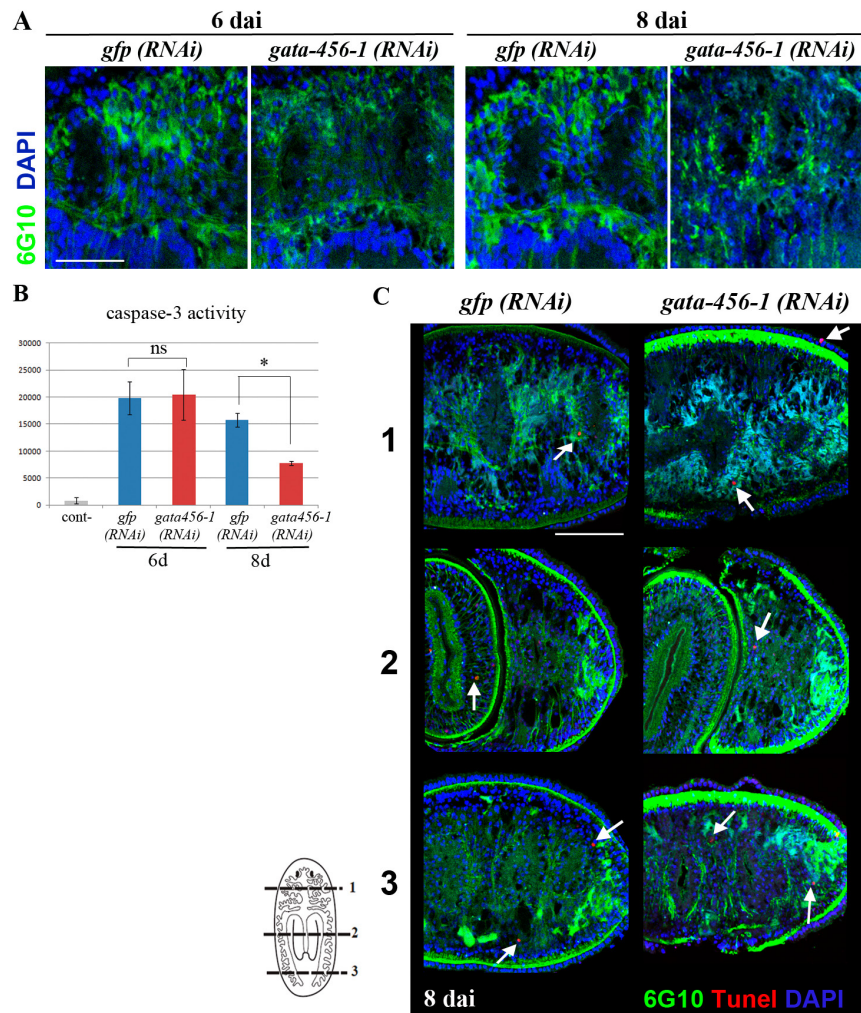
Suppl. Fig. 5. Gut cells disappearance after *Smed-gata456-1(RNAi)* in regenerating animals. Immunohistochemistry using the muscular antibody 6G10 (green) combined with nuclear DAPI staining (blue) in transverse sections of a regenerating tail at 8 days of regeneration. The images correspond to the ones shown in Fig. 1G. Scale bar: 100 μ m.



Suppl. Fig. 6. Anterior mediolateral patterning is impaired in regenerating *Smed-gata456-2(RNAi)* animals. (A) Live regenerating control and *Smed-gata456-2(RNAi)* animals after 12 days of regeneration. (B) Immunohistochemistry with the eye marker anti-arrestin. (C) Whole mount in situ hybridization for the anterior marker *notum*. Arrowheads indicate lateral-most *notum*-expressing cells, which collapse in the midline of *Smed-gata456-2(RNAi)* animals. (D) Whole mount in situ hybridization for the brain branch marker *gas*, showing the fusion of the two cephalic ganglia in *Smed-gata456-2(RNAi)* animals. (E) Quantification of mitotic cells by anti-H3P immunohistochemistry in the postblastema after 6 hours ($n=5$ controls and $n=5$ *Smed-gata456-2(RNAi)* animals) and 48 hours ($n=5$ controls and $n=4$ *Smed-gata456-2(RNAi)* animals) of regeneration. Error bars represent the standard error of the mean. Data were analyzed by Student's *t*-test. Differences are considered significant at $P<0.05$. (F) Whole mount fluorescent in situ hybridization for the intestinal marker *pk-1*, showing the anterior region of a regenerating trunk. (G) Whole mount fluorescent in situ hybridization for the intestinal marker *pk-1*, showing the posterior region of a regenerating trunk. (H) Quantification of the relative length of the anterior gut branch with respect to the prepharyngeal region. Error bars represent the standard error of the mean. Data were analyzed by Student's *t*-test. $*P<0.05$. Scale bars: 500 μm in A; 50 μm in B; 150 μm in C-D; 200 μm in F-G. Experiments were performed in regenerating trunks after 12 days of regeneration, following 2 weeks of injection and regeneration.



Suppl. Fig. 7. *Smed-gata456-1(RNAi)* phenotype in intact animals. (A) Live intact control and *Smed-gata456-1(RNAi)* animals, 12 days after the last dsRNA injection. **(B)** Graph showing survival of *Smed-gata456-1(RNAi)* intact animals ($n=52$). Red arrowheads indicate lesions. dai, days after the last injection. The anterior end is oriented to the left. **(C)** Immunohistochemistry using the muscular antibody 6G10 (green) combined with nuclear DAPI staining (blue) in transverse sections of an intact animal 12 days after the last injection. The images correspond to the ones shown in Fig. 3A. dai, days after injection. Scale bars: 500 μm in A; 100 μm in C.



Suppl. Fig. 8. Apoptosis is unaffected in *Smed-gata456-1*(RNAi) animals. (A) Immunohistochemistry using the muscular antibody 6G10 (green) combined with nuclear DAPI staining (blue) in transverse sections of the head region of an intact animal 6 and 8 days after the last dsRNA injection. The clear disruption of the gastrodermis at 8 days but not at 6 days after the last dsRNA injection indicates that at this time point the gut phenotype is set up. **(B)** Quantification of caspase-3 activity in intact animals 6 and 8 days after the last dsRNA injection ($n=3$). Error bars represent the standard error of the mean. Data were analyzed by Student's *t*-test. * $P<0.05$; differences are considered significant at $P<0.05$. **(C)** TUNEL staining in transverse sections corresponding to pre-pharyngeal, pharyngeal and tail regions of intact animals 8 days after the last dsRNA injection. Sections are also immunostained with the muscular antibody 6G10 (green) and DAPI (blue). Arrows point to TUNEL + cells. Scale bars: 50 μm in A; 100 μm in C.