modERN omnibus Gdoc

- Dmel sheets: marked by red
  - modENCODE_Dmel
    - Still need to populate with data from my spreadsheet
  - modERN_Dmel
  - modERN_Dmel_post_freeze
  - Dmel_controls
  - Dmel_post_processing
    - Information on where to get other items eg: HOTregions

- Cele: marked in blue
  - Equivalent sheets to Dmel

Omnibus modERN Dmel Sheet

2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	FlyBase Symbol	IGSBmod Factor Name	isoform	MiMIC	BDSC stock ID/DCC ID	Short Stage	Stage	sex	age	Source	Antibody ID	Rep	BID/file	igsbmod location	DCC link	DCC status	dm3 Uniq. Map. Reads	dm3 IDR peaks	dm3 N1/N2	dm3 Np/Nt	wt_subtracted IDR peaks	wt_subtracted folde
756	Xbp1	xbp1			ENCDO555HC	E0-24	embryonic		0-24	Input	INPUT	INPUT	2016-673	/raid/modencode/dmel/processed/dm3/20160702	https://www.encodeproject.org/ENCSR698IEF	released	17,792,402	6939	1.43	1.21		
757	rgr	rgr			65289	W3L	wandering third instar larva			IP	GoatV-aGFP	1	2016-674	/raid/modencode/dmel/processed/dm3/20160702	https://www.encodeproject.org/ENCSR951FDL	released	8,703,745	8497	1.67	1.03		
758	rgr	rgr			65289	W3L	wandering third instar larva			IP	GoatV-aGFP	2	2016-675	/raid/modencode/dmel/processed/dm3/20160702	https://www.encodeproject.org/ENCSR951FDL	released	13,618,950	8497	1.67	1.03		
759	rgr	rgr			65289	W3L	wandering third instar larva			IP	GoatV-aGFP	3	2016-676	/raid/modencode/dmel/processed/dm3/20160702	https://www.encodeproject.org/ENCSR951FDL	released	7,333,840	8497	1.67	1.03		
760	rgr	rgr			65289	W3L	wandering third instar larva			Input	INPUT	INPUT	2016-677	/raid/modencode/dmel/processed/dm3/20160702	https://www.encodeproject.org/ENCSR951FDL	released	17,812,779	8497	1.67	1.03		
761	Clamp	CG1832			ENCDO839RF	E0-24	embryonic		0-24	IP	GoatV-aGFP	2	2016-805	/raid/modencode/dmel/processed/dm3/20160819	https://www.encodeproject.org/ENCSR858HYS	proposed	11,262,257	3025	1.26	1.81		
762	Clamp	CG1832			ENCDO839RF	E0-24	embryonic		0-24	IP	GoatV-aGFP	4	2016-918	/raid/modencode/dmel/processed/dm3/20160819	https://www.encodeproject.org/ENCSR858HYS	proposed	13,093,994	3025	1.26	1.81		
763	Clamp	CG1832			ENCDO839RF	E0-24	embryonic		0-24	Input	NA	INPUT	2016-807	/raid/modencode/dmel/processed/dm3/20160819	https://www.encodeproject.org/ENCSR858HYS	proposed	6,730,381	3025	1.26	1.81		
764	tio	tio			38671	E4-24	embryonic		4-24	IP	GoatV-aGFP	1	2016-779	/raid/modencode/dmel/processed/dm3/20160819	https://www.encodeproject.org/ENCSR748DHJ	proposed	17,359,628	3740	1.02	1.10		
765	tio	tio			38671	E4-24	embryonic		4-24	IP	GoatV-aGFP	2	2016-780	/raid/modencode/dmel/processed/dm3/20160819	https://www.encodeproject.org/ENCSR748DHJ	proposed	14,691,179	3740	1.02	1.10		
766	tio	tio			38671	E4-24	embryonic		4-24	Input	NA	INPUT	2016-782	/raid/modencode/dmel/processed/dm3/20160819	https://www.encodeproject.org/ENCSR748DHJ	proposed	24,377,425	3740	1.02	1.10		
767	Su(var)2-10	Suvar2-10			64795	E0-24	embryonic		0-24	IP	GoatV-aGFP	2	2016-1573	/raid/modencode/dmel/processed/dm3/20160819			14,707,701	2745	1.07	1.06		
768	Su(var)2-10	Suvar2-10			64795	E0-24	embryonic		0-24	IP	GoatV-aGFP	3	2016-1574	/raid/modencode/dmel/processed/dm3/20160819			13,928,478	2745	1.07	1.06		
769	Su(var)2-10	Suvar2-10			64795	E0-24	embryonic		0-24	Input	NA	INPUT	2016-1575	/raid/modencode/dmel/processed/dm3/20160819			15,790,199	2745	1.07	1.06		
770	pita	Pita			64792	E0-24	embryonic		0-24	IP	GoatV-aGFP	1	2016-1584	/raid/modencode/dmel/processed/dm3/20160819			9,218,404	1319	1.03	1.14		
771	pita	Pita			64792	E0-24	embryonic		0-24	IP	GoatV-aGFP	3	2016-1586	/raid/modencode/dmel/processed/dm3/20160819			10,262,464	1319	1.03	1.14		
772	pita	Pita			64792	E0-24	embryonic		0-24	Input	NA	INPUT	2016-1587	/raid/modencode/dmel/processed/dm3/20160819			11,510,717	1319	1.03	1.14		
773	Clk	Clk			64789	AM	adult	male		IP	GoatV-aGFP	1	2016-1629	/raid/modencode/dmel/processed/dm3/20160819			8,928,705	2344	1.43	1.11		
774	Clk	Clk			64789	AM	adult	male		IP	GoatV-aGFP	2	2016-1630	/raid/modencode/dmel/processed/dm3/20160819			5,519,870	2344	1.43	1.11		
775	Clk	Clk			64789	AM	adult	male		Input	NA	INPUT	2016-1632	/raid/modencode/dmel/processed/dm3/20160819			12,124,738	2344	1.43	1.11		
776	org-1	org-1			64786	E2-24	embryonic		2-24	IP	GoatV-aGFP	1	2016-1614	/raid/modencode/dmel/processed/dm3/20160819			12,894,389	4012	1.27	1.57		
777	org-1	org-1			64786	E2-24	embryonic		2-24	IP	GoatV-aGFP	2	2016-1615	/raid/modencode/dmel/processed/dm3/20160819			9,996,832	4012	1.27	1.57		
778	org-1	org-1			64786	E2-24	embryonic		2-24	Input	NA	INPUT	2016-1617	/raid/modencode/dmel/processed/dm3/20160819			21,037,921	4012	1.27	1.57		
779	CG32006	CG32006			59982	WPP+48	prepupa		48	IP	GoatV-aGFP	1	2016-1724	/raid/modencode/dmel/processed/dm3/20160830			14,252,817	11625	1.30	1.10		
780	CG32006	CG32006			59982	WPP+48	prepupa		48	IP	GoatV-aGFP	2	2016-1725	/raid/modencode/dmel/processed/dm3/20160830			26,275,617	11625	1.30	1.10		
781	CG32006	CG32006			59982	WPP+48	prepupa		48	IP	GoatV-aGFP	3	2016-1726	/raid/modencode/dmel/processed/dm3/20160830			13,935,152	11625	1.30	1.10		
782	CG32006	CG32006			59982	WPP+48	prepupa		48	Input	INPUT	INPUT	2016-1727	/raid/modencode/dmel/processed/dm3/20160830			18,585,139	11625	1.30	1.10		
783	Meics	Meics			64796	E0-24	embryonic		0-24	IP	GoatV-aGFP	1	2016-1740	/raid/modencode/dmel/processed/dm3/20160830			16,148,256	7313	1.34	1.02	1753	/raid/modencode/d
784	Meics	Meics			64796	E0-24	embryonic		0-24	IP	GoatV-aGFP	3	2016-1742	/raid/modencode/dmel/processed/dm3/20160830			17,364,783	7313	1.34	1.02	1753	/raid/modencode/d
785	Meics	Meics			64796	E0-24	embryonic		0-24	Input	INPUT	INPUT	2016-1743	/raid/modencode/dmel/processed/dm3/20160830			15,765,938	7313	1.34	1.02	1753	/raid/modencode/d
786	bcd	bcd			30869	E0-24	embryonic		0-24	IP	GoatV-aGFP	1	2016-1736	/raid/modencode/dmel/processed/dm3/20160830			13,824,627	5787	1.05	1.40	145	/raid/modencode/d
787	bcd	bcd			30869	E0-24	embryonic		0-24	IP	GoatV-aGFP	3	2016-1738	/raid/modencode/dmel/processed/dm3/20160830			20,268,285	5787	1.05	1.40	145	/raid/modencode/d

1

3

4

5

1: typical metadata
2: includes stock reference
3: location on IGSBmod server

4: DCC info, need to make link
5: Processing results for both genomes

Omnibus modENCODE Cele Sheet

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Factor Name	Strain/ OP#	Stage	Source	Antibody ID	Rep	BID/file	Notes	igsbmod location 3	modENCODE. org ID	GEO ID 4	DCC link	DCC status	WS220 Mapping Reads	WS220 IDR peaks	WS220 N1/N2	6 WS220 Np/Nt	WS235 Mapping Reads
942	PHA-4	OP37	LE	INPUT	NA	2	SRR107552.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161004	2598	GSM677657	ENCSR304RME	Ready for release	19636	21	1.125	21	19636
943	PHA-4	OP37	LE	IP	GFP	2	SRR107553.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161004	2598	GSM677658	ENCSR304RME	Ready for release	19636	21	1.125	21	19636
944	PHA-4	OP37	L4	INPUT	NA	1	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160106	2945	GSM1183581	ENCSR494DGU	Ready for release	8136625	1622	2.076923077	1.110198494	8138411
945	PHA-4	OP37	L4	INPUT	NA	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160106	2945	GSM1183583	ENCSR494DGU	Ready for release		1622	2.076923077	1.110198494	
946	PHA-4	OP37	L4	IP	GFP	1	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160106	2945	GSM1183582	ENCSR431IBS	Ready for release	720412	1622	2.076923077	1.110198494	720597
947	PHA-4	OP37	L4 1 IP	GFP	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160106	2945	GSM1183584	ENCSR431IBS	Ready for release	684080	1622	2.076923077	1.110198494	684252	
948	PHA-4	OP37	L2	INPUT	NA	1	SRR164264.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	3158	GSM700181	DCC missing replicates		31902	148	1.243243243	1.129770992	31912
949	PHA-4	OP37	L2	INPUT	NA	1	SRR164263.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	3158	GSM700180	DCC missing replicates		31902	148	1.243243243	1.129770992	31912
950	PHA-4	OP37	L2	IP	GFP	1	SRR164262.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	3158	GSM700179	DCC missing replicates		14259	148	1.243243243	1.129770992	14268
951	PHA-4	OP37	L2	IP	GFP	2	SRR164261.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	3158	GSM700178	DCC missing replicates	5	22401	148	1.243243243	1.129770992	22405
952	PHA-4	OP37	L2	IP	GFP	2	SRR164260.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	3158	GSM700177	DCC missing replicates		22401	148	1.243243243	1.129770992	22405
953	PHA-4	OP37	YA	INPUT	NA	1	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3161	GSM700193	_		9782056	923	0.01930215293	102.5555556	9784109
954	PHA-4	OP37	YA	INPUT	NA	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3161	GSM700191	_			923	0.01930215293	102.5555556	
955	PHA-4	OP37	YA	INPUT	NA	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3161	GSM700192	_			923	0.01930215293	102.5555556	
956	PHA-4	OP37	YA	IP	GFP	1	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3161	GSM700197			2947778	923	0.01930215293	102.5555556	2948433
957	PHA-4	OP37	YA	IP	GFP	1		2 Not used for processing, not included in Lou's info		3161	GSM700196	didn't process with this						
958	PHA-4	OP37	YA	IP	GFP	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3161	GSM700194			3078841	923	0.01930215293	102.5555556	3079574
959	PHA-4	OP37	YA	IP	GFP	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3161	GSM700195			3078841	923	0.01930215293	102.5555556	3079574
960	PHA-4	OP37	L3	INPUT	NA	1	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3215	GSM729296	ENCSR415CLB	Ready for release	13369315	7854	1.248754063	1.145900204	13372218
961	PHA-4	OP37	L3	INPUT	NA	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3215	GSM729297	ENCSR415CLB	Ready for release		7854	1.248754063	1.145900204	
962	PHA-4	OP37	L3	INPUT	NA	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3215		ENCSR415CLB	Ready for release		7854	1.248754063	1.145900204	
963	PHA-4	OP37	L3	IP	GFP	1	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3215	GSM729298	ENCSR845QMZ	Ready for release	1943627	7854	1.248754063	1.145900204	1944152
964	PHA-4	OP37	L3	IP	GFP	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3215	GSM729299	ENCSR845QMZ	Ready for release	6103612	7854	1.248754063	1.145900204	6105140
965	PHA-4	OP37	L3	IP	GFP	2	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160118	3215		ENCSR845QMZ	Ready for release	6103612	7854	1.248754063	1.145900204	6105140
966	PHA-4	OP37	L4	INPUT	NA	1	SRR947493.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	4033	GSM1183785	_		5510333	1599	1.439622642	0.5922451532	5511501
967	PHA-4	OP37	L4	INPUT	NA	2	SRR947495.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	4033	GSM1183787	_		5746132	1599	1.439622642	0.5922451532	5747413
968	PHA-4	OP37	L4	INPUT	NA	3	SRR947497.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	4033	GSM1183789	DCC missing replicates		5642681	1599	1.439622642	0.5922451532	5643981
969	PHA-4	OP37	L4	IP	GFP	1	SRR947494.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	4033	GSM1183786	_		4280513	1599	1.439622642	0.5922451532	4281482
970	PHA-4	OP37	L4	IP	GFP	2	SRR947496.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	4033	GSM1183788	_		2412430	1599	1.439622642	0.5922451532	2413040
971	PHA-4	OP37	L4	IP	GFP	3	SRR947498.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20161029	4033	GSM1183790	DCC missing replicates		2409842	1599	1.439622642	0.5922451532	2410421
972	PQM-1	OP201	L3	INPUT	NA	3	http://encodedcc.stanford.edu/ftp/modENCODE_VS_ENCO		/raid/modencode/ce/processed/WS220/old_data/20160127	2623, Unknown	GSM677695	ENCSR042QHV	Ready for release	0	5666	3.388444444	1.561742007	0
973	PQM-1	OP201	L3	INPUT	NA	5	http://encoded...T_pf.fastq.gz		/raid/modencode/ce/processed/WS220/old_data/20160127	2623, Unknown	GSM677697	ENCSR042QHV	Ready for release	0	5666	3.388444444	1.561742007	0

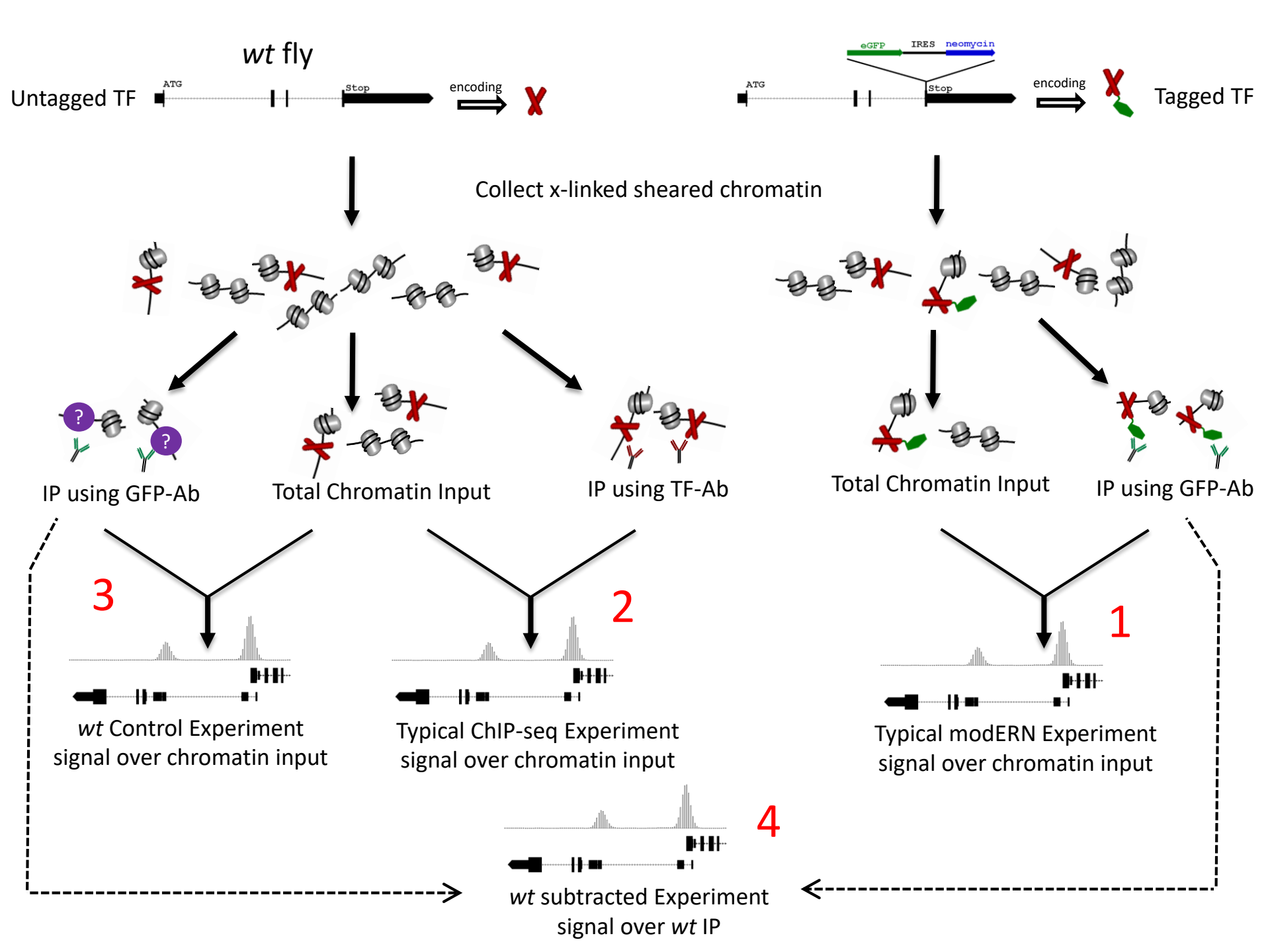
1: typical metadata
2: file used, or not
3: location on IGSBmod server

4: DB info, old and new
5: Still working on uploading
6: run data for both genomes

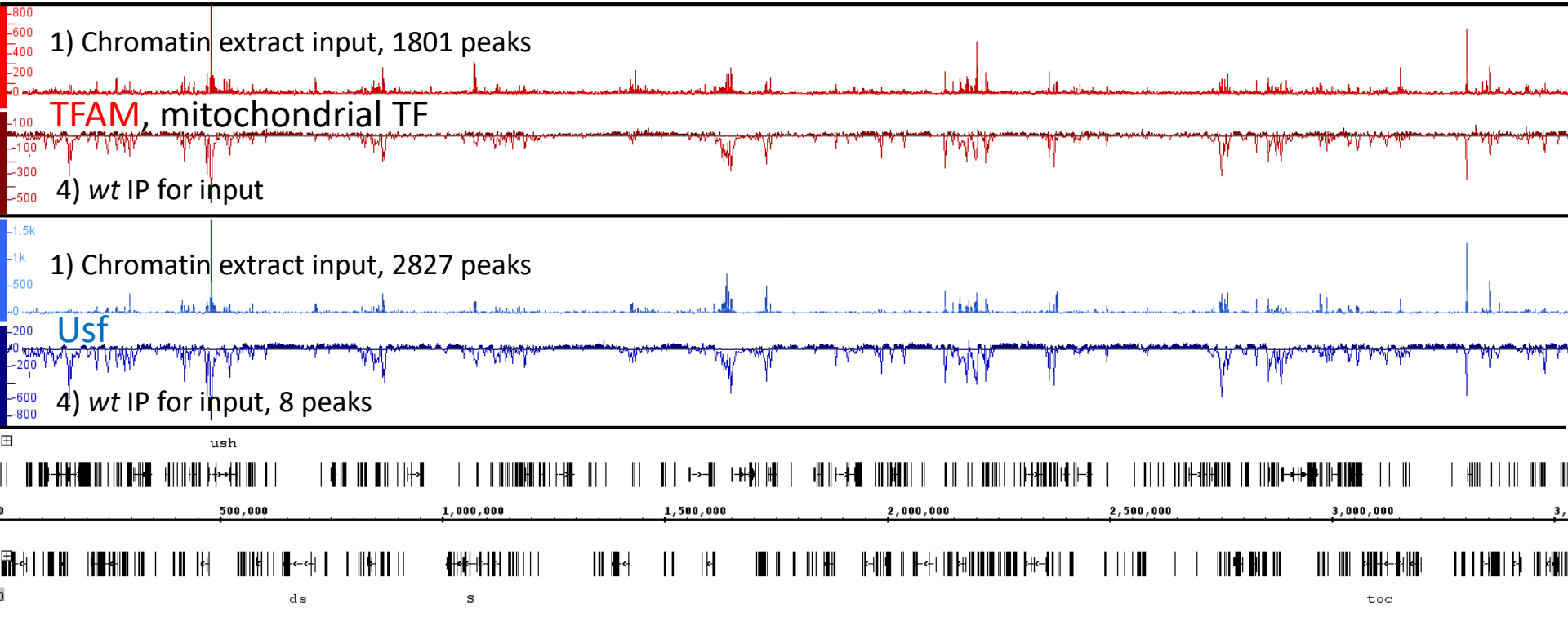
Need to split into sheet with Lou's list and sheet with *et alia*

Progress Summary

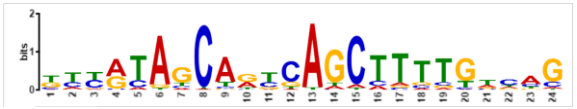
- Dmel:
  - Old Dmel data run, need to double-check datasets and files used
  - Some modENCODE fastq's uploaded to DCC
    - Need to finish.
  - ~25 new modERN datasets need to be uploaded
  - ~4 modERN datasets awaiting release for Ab290 vs GoatV comparison
  - 17 sets awaiting sequencing
- Cele:
  - See if "non-Lou" data is good, decide on uploading (80 datasets not registered by Esther, but some might be on the list)
  - Create new DCC replicate objects to finish uploading old data (~20)
  - 3 new datasets not loaded
  - Ab comparison for 20 sets to be released
  - 12 sets in pipeline at UofC


wt fly
Untagged TF
ATG
Stop
encoding
eGFP
IRES
neomycin
ATG
Stop
encoding
Tagged TF
Collect x-linked sheared chromatin
IP using GFP-Ab
Total Chromatin Input
IP using TF-Ab
Total Chromatin Input
IP using GFP-Ab
3
2
1
wt Control Experiment
signal over chromatin input
Typical ChIP-seq Experiment
signal over chromatin input
Typical modERN Experiment
signal over chromatin input
4
wt subtracted Experiment
signal over wt IP

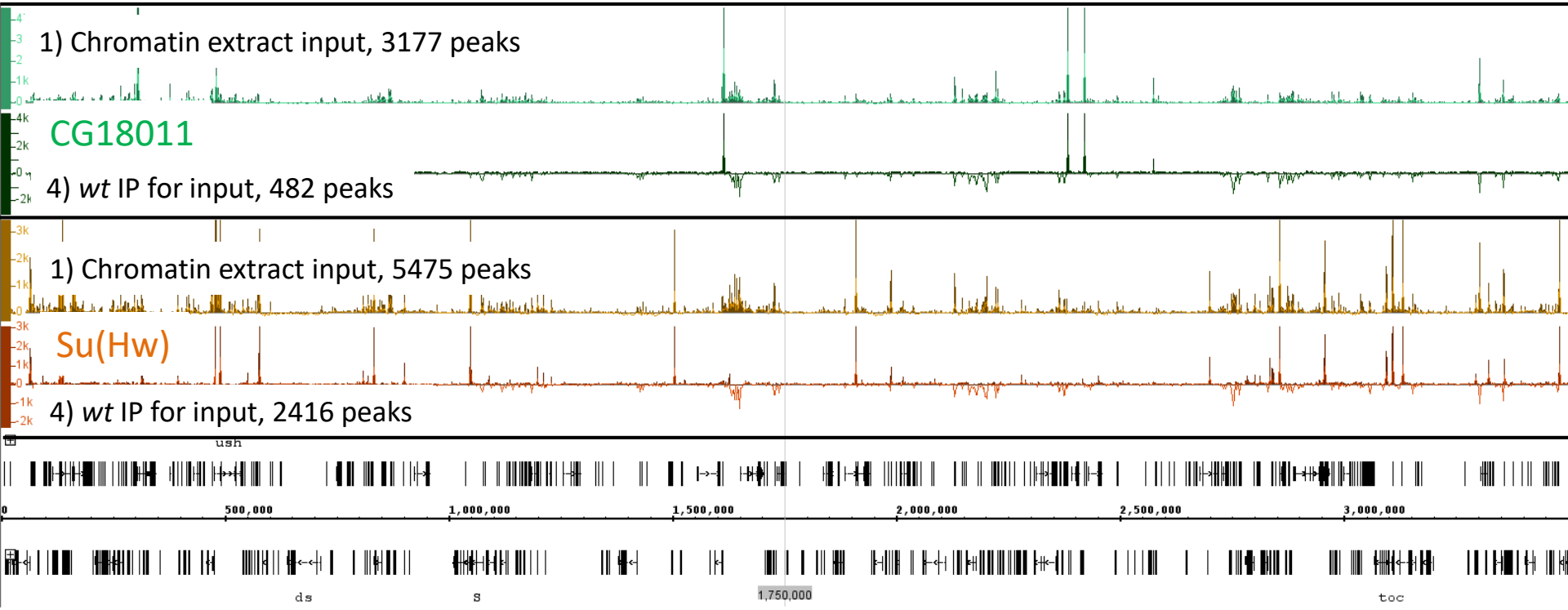
2 Questionable Datasets

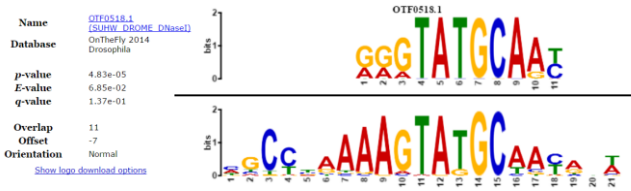


2 Good Datasets



MEME predicted CG18011(C2H2 ZF) motif with top 50 peaks E=2.4e-069





Reported Su(Hw) motif

MEME predicted Su(Hw) motif with top 50 peaks
E=2.3e-119

Final items

- Tried comparison between TF-Ab vs GFP-Ab, both wt-IP subtracted.
  - Motifs in GFP matched expected motifs when available
  - TF-Ab was new IPs not from modENCODE, matched wrong genes, and ¾ matched phol/PRE motif
- Plan on trying to subtract wt-IP datasets generated with Abcam GFP-Ab and FLAG-Ab from TF data.
  - Both have >50% overlap in peaks with GoatV.
  - Same peaks remain compared to GoatV.
- Expand MiMIC lines with GFP-FLAG tag.
  - Will same peaks remain when wt-FLAG-IP is subtracted from TF-FLAG IP?