Copy number variation and evolution in humans and chimpanzees

Perry GH *et al.*

Genome Res. 18:1698- 1697

Journal Club by Tarmo Puurand

Platform

- Array-based comparative genomic hybridisation (aCGH) on human whole-genome tile-path (WGTP)
- 28708 large-insert DNA clones (upgraded version contains 2000 additional clones compared to previous generation of the WGTP array)

Samples

- 30 chimpanzees, 3 are not wild-born, 29 Western chimanzee, 1 Eastern chimp
- 30 human samples
- 10 Yoruba, Nigeria
- 10 Biaka rainforest, Central African Republic
- 10 Mbuti rainforest, Democratic Republic of Congo

References for aCGH

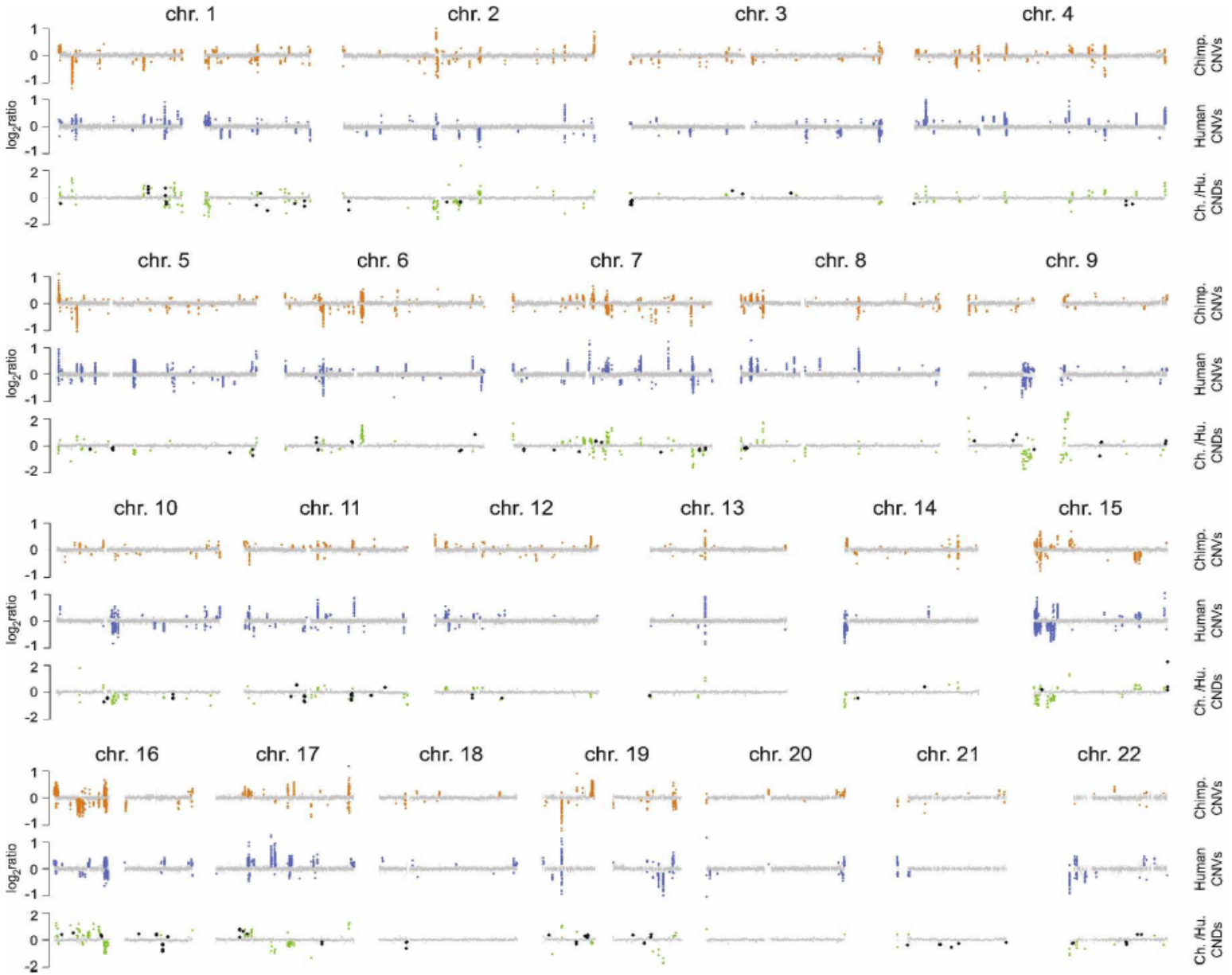
- For humans
  - Europan-American male NA10851
- For chimpanzees
  - Clint- the captive-born for the chimpanzee reference genome sequence

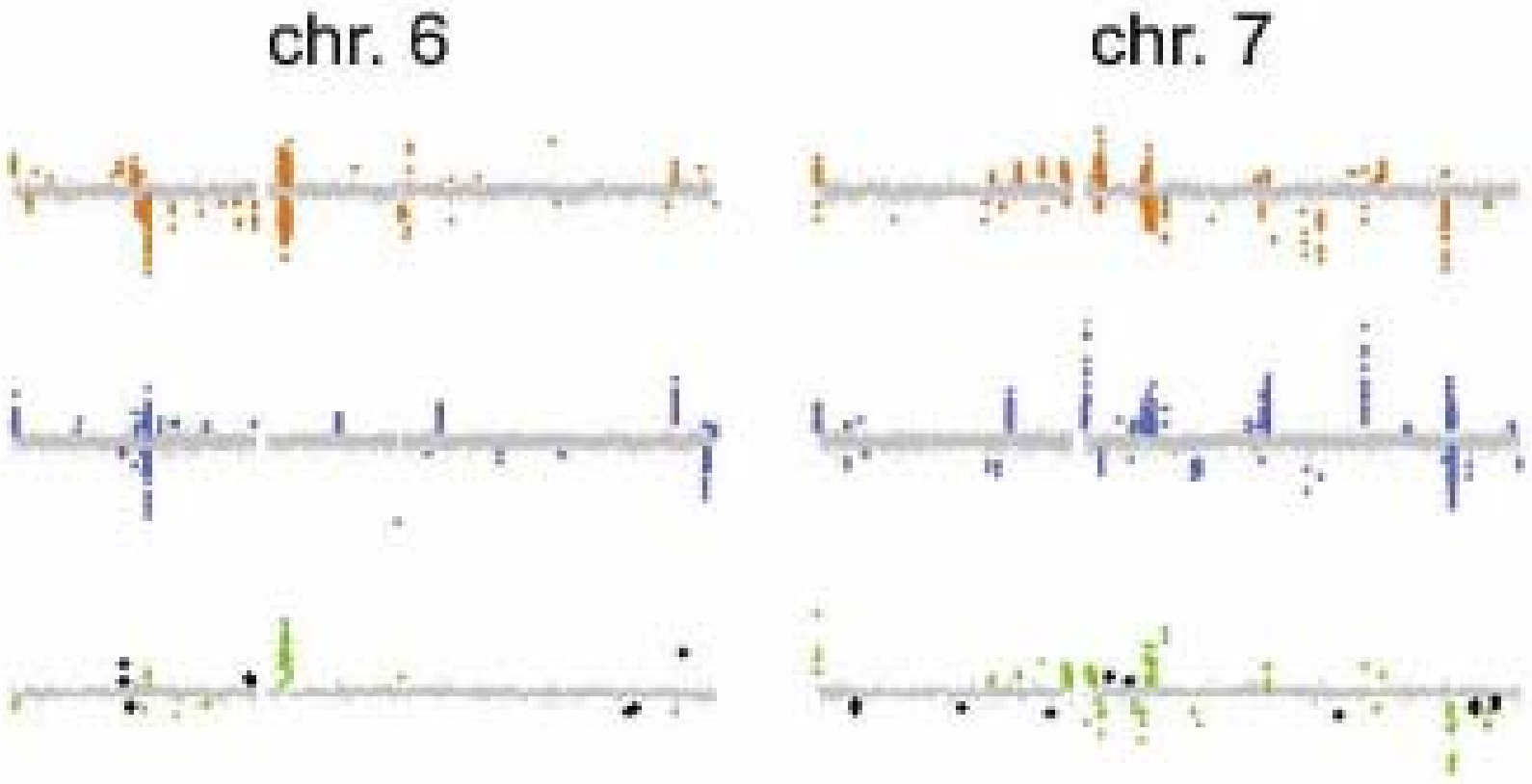
Detection of CNV differences

- CNVfinder, enables to find CNV losses and gains on the WGTP platform with a false positive rate < 5%
- Comparision with human CNVs in the Database of Genomic Variation
- FISH

Results

- 70 and 80 autosomal CNVs per within-chimpanzee and within-human comparision
- Median size ~250 kb for both species
- 353 discrete autosomal CNV-containing regions (CNVRs) in humans
- 438 CNVRs in chimpanzees


chr. 1
chr. 2
chr. 3
chr. 4
chr. 5
chr. 6
chr. 7
chr. 8
chr. 9
chr. 10
chr. 11
chr. 12
chr. 13
chr. 14
chr. 15
chr. 16
chr. 17
chr. 18
chr. 19
chr. 20
chr. 21
chr. 22
log2ratio
Chimp. CNVs
Human CNVs
Ch./Hu. CNDs


chr. 6
chr. 7

Results

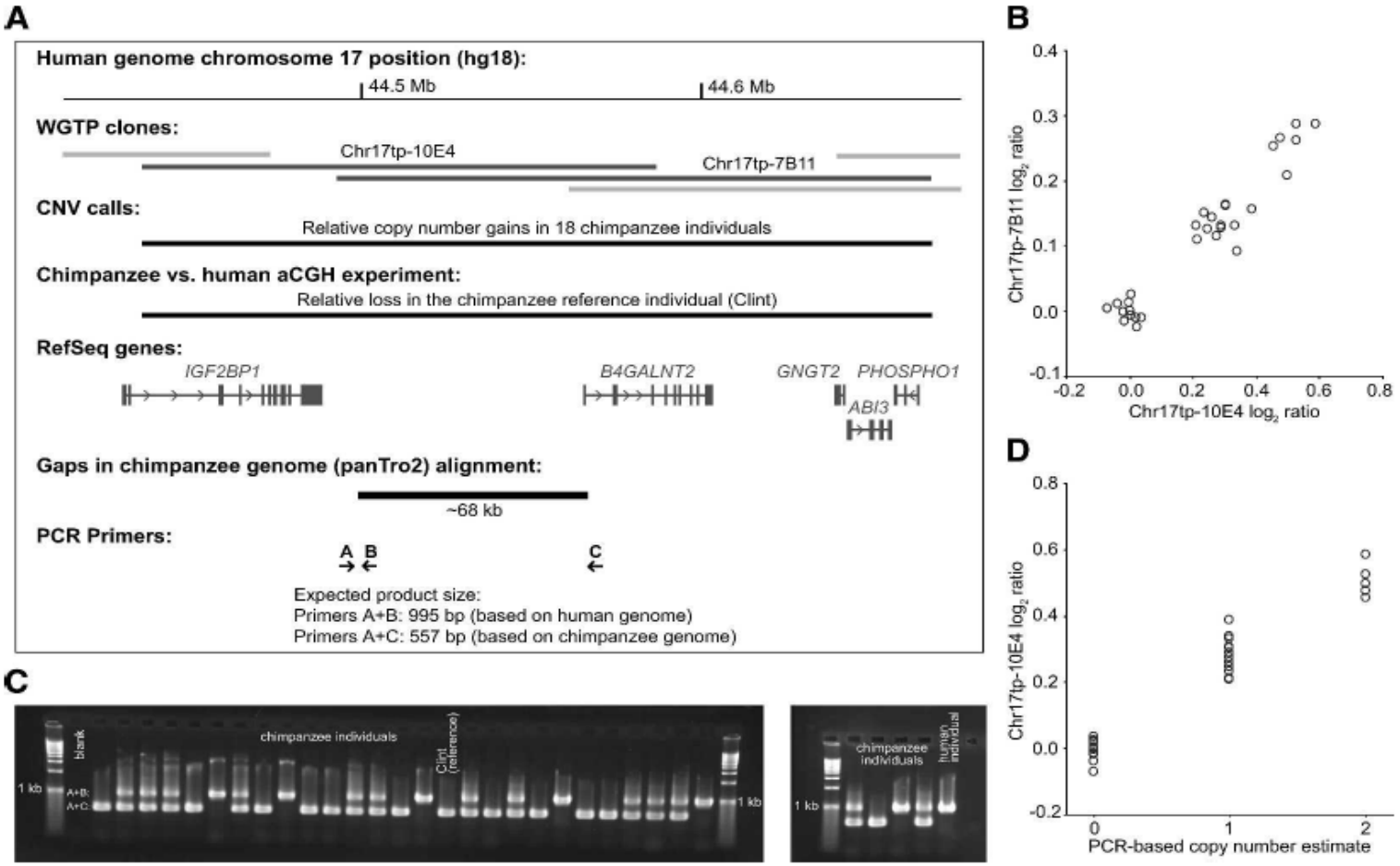
- 313 of the 353 (89%) overlap with similar previous human studies. 222 (63%) were expected to overlap at random based on permutation analysis.
- Only 53 of 438 (12,1%) overlap with similar chimpanzee CNVR studies (two aCGH platforms with ~12% genomic coverage)

CNDs

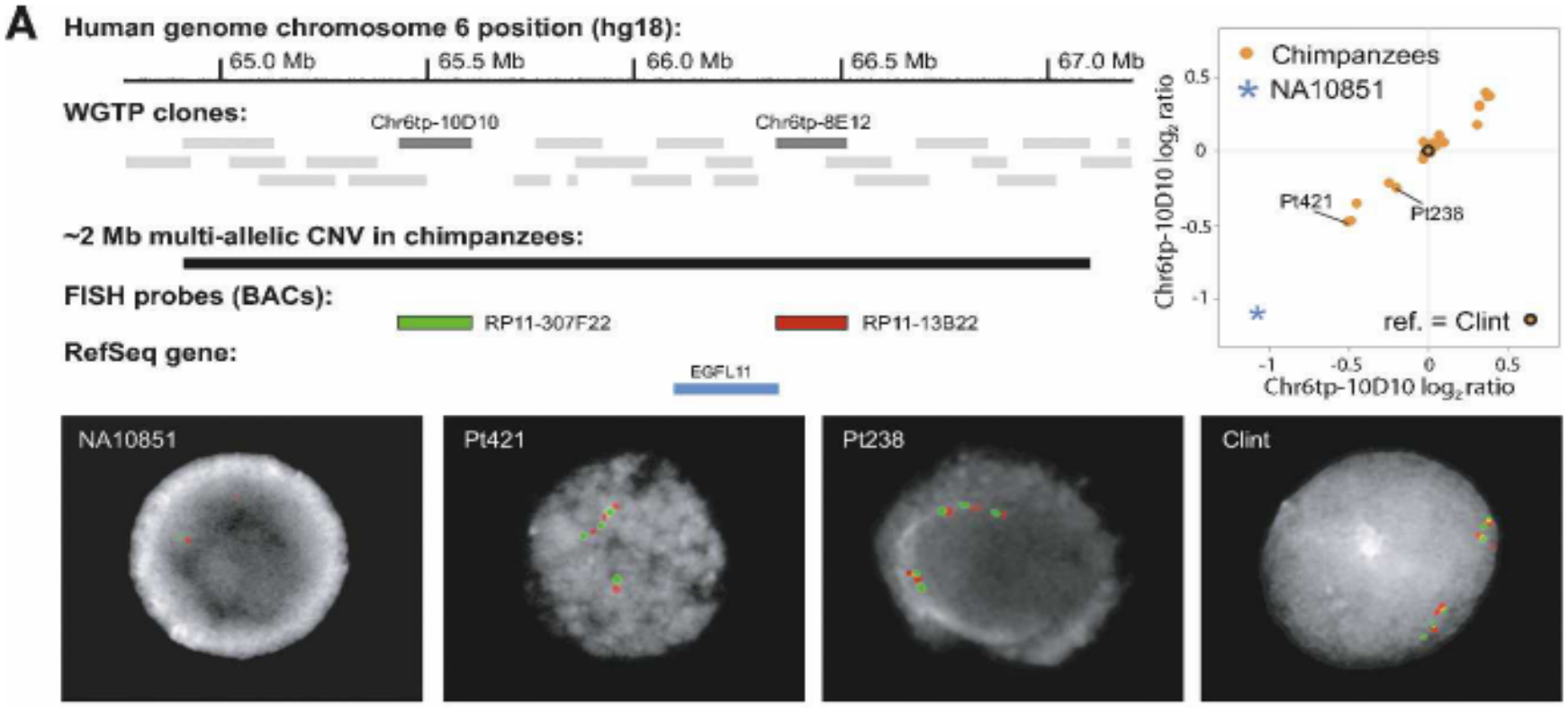
- 355 between-species copy number differences (CNDs) were identified
- 140 of them overlapped with previously identified CNDs

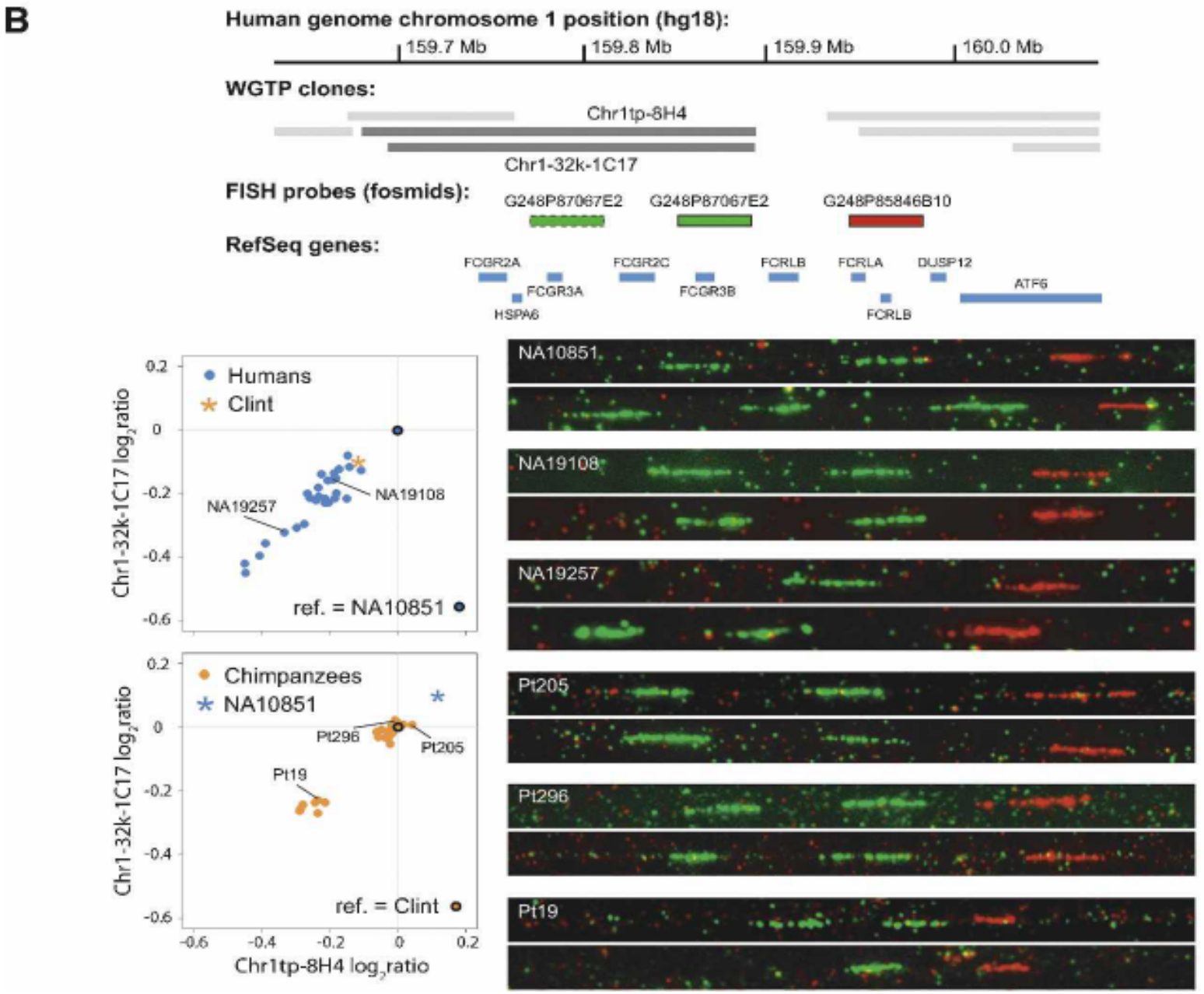
- Among the 438 observed chimp CNVRs, 9 putative deletion variants in regions with gaps in Clint’s genome were found

CNV deletion in chimp



FISH validation




B
Human genome chromosome 1 position (hg18):
159.7 Mb
159.8 Mb
159.9 Mb
160.0 Mb
WGTP clones:
Chr1tp-8H4
Chr1-32k-1C17
FISH probes (fosmids):
G248P87067E2
G248P87067E2
G248P85846B10
RefSeq genes:
FCGR2A
FCGR3A
HSPA6
FCGR2C
FCGR3B
FCRLB
FCRLA
FCRLB
DUSP12
ATF6
Humans
Clint
NA19257
NA19108
ref. = NA10851
Chr1-32k-1C17 log2ratio
Chimpanzees
NA10851
Pt296
Pt205
Pt19
ref. = Clint
Chr1tp-8H4 log2ratio
NA10851
NA19108
NA19257
Pt205
Pt296
Pt19

Results

- 144 of the 353 human CNVRs (42%) overlapped with chimpanzee CNVRs (random expectation 39, 11%)
- 182 of the 353 human CNVRs (51,6%) overlapped with segmental duplication (SDs) regions in the human genome (random expectation 47, 13,4%)
- Similar level of enrichment was observed for chimpanzees

CNV frequency distribution analysis

Table 2. Gene contents and frequency distributions of human and chimpanzee CNVRs

Genes		Human CNVRs				Chimpanzee CNVRs			
Category[a]	Description	R[b]	C[b]	Ratio R/C	Score[c]	R[b]	C[b]	Ratio R/C	Score[c]
All genes	—	40	137	0.29	1.00	57	121	0.47	1.00
Lowest scores									
GO:0006952	Defense response	0	13	0.00	0.07	0	9	0.00	0.10
GO:0006629	Lipid metabolic process	0	10	0.00	0.09	0	7	0.00	0.13
GO:0003924	GTPase activity	0	11	0.00	0.08	2	8	0.25	0.58
GO:0006886	Intracellular protein transport	0	12	0.00	0.08	2	6	0.33	0.75
GO:0004871	Signal transducer activity	3	10	0.30	1.03	0	12	0.00	0.08
GO:0004984	Olfactory receptor activity	3	10	0.30	1.03	0	9	0.00	0.10
Highest Scores									
GO:0007601	Visual perception	3	7	0.43	1.41	3	4	0.75	1.47
GO:0005488	Binding	4	9	0.44	1.47	4	5	0.80	1.58
GO:0004674	Protein serine/threonine kinase activity	4	8	0.50	1.63	5	7	0.71	1.45
GO:0043565	Sequence-specific DNA binding	5	11	0.45	1.51	6	5	1.20	2.29
GO:0004725	Protein tyrosine phosphatase activity	4	6	0.67	2.10	2	2	1.00	1.75
GO:0006470	Protein amino acid dephosphorylation	5	6	0.83	2.59	2	2	1.00	1.75

[a]Selected Gene Ontology (GO) categories, with ≥10 SD-containing CNVRs that overlap one or more genes of a given GO category in at least one species.
[b]R, rare (frequency = 1); C, common (frequency ≥ 2).
[c]The score is a normalized R/C ratio for each GO category. It was calculated for each species using the formula (1 + R/A)/(1 + C), where A is the ratio R/C for all genes. Only the GO categories with the six lowest and six highest averaged scores are listed.

Comparision of copy number fixation and polymorphism

Table 3. Rates of copy number fixation and polymorphism by gene functional categories

GO categories[a]	Description	Fixed CNDs[b]	Total CNVRs[c]	Ratio *F*/*T*	Score[d]	*P*-value[e]
—	No gene (intergenic)	18	117	0.15	1.00	NA
—	One or more gene(s)	74	518	0.14	0.93	0.886
Lowest scores						
GO:0008233	Peptidase activity	0	25	0.00	0.04	0.048
GO:0048503	GPI anchor binding	0	23	0.00	0.04	0.077
GO:0016301	Kinase activity	0	18	0.00	0.05	0.132
GO:0006811	Ion transport	1	48	0.02	0.15	0.027
GO:0005215	Transporter activity	1	45	0.02	0.16	0.029
Other scores (discussed in text)						
GO:0006955	Immune response	3	35	0.09	0.57	0.420
GO:0004984	Olfactory receptor activity	5	20	0.25	1.60	0.534
Highest scores						
GO:0005506	Iron ion binding	8	28	0.29	1.83	0.197
GO:0051301	Cell division	5	15	0.33	2.09	0.182
GO:0007067	Mitosis	5	15	0.33	2.09	0.182
GO:0008283	Cell proliferation	6	15	0.40	2.50	0.099
GO:0006954	Inflammatory response	5	12	0.42	2.58	0.141

[a]Gene Ontology (GO) categories were included in the analysis only if $F + T > 16$, where F is the number of fixed CNDs and T the total number of CNVRs with one or more genes from the GO category.
[b]The number of CNDs between the human and chimpanzee reference individuals that did not overlap any within-species human or chimpanzee CNVR, that overlap one or more genes assigned to a given GO category.
[c]The number of total CNVRs (human-only CNVRs + chimpanzee-only CNVRs + CNVRs observed in the same regions in both species; i.e., no CNVR regions are counted twice) that overlap one or more genes assigned to a given GO category.
[d]The score is a normalized F/T ratio for each GO category. It was calculated using the formula $(1 + F/A)/(1 + T)$, where A is the ratio F/T for all CNDs/CNVRs that do not contain genes (intergenic variants). The GO categories with the five lowest and five highest scores are listed, as well as two categories discussed in the text: see Supplemental Table 4 for complete data set.
[e]Two-tailed Fisher's exact tests for each GO category versus the intergenic F/T ratio (CNDs/CNVRs). *P*-values are not corrected for multiple tests.